

Negar Nejatishahidin

I have used argo instead of Google Colab. So [for the training part](#) I use the Main.py file, which is available in my HW folder, instead of Ipython file. I have an [out.txt](#) file which I wright all the outputs there. I have also copy them here.

[For the test part](#), I used Google Cola and [I have attached the pictures at the end.](#)

```
Load pre-trained model
Initializing dataset
Loaded 5011 train images
Initializing dataset
Loaded 4950 test images
```

```
Starting epoch 1 / 50
Learning Rate for this epoch: 0.001
Epoch [1/50], Iter [5/209] Loss: 59.9375, average_loss: 71.4931
Epoch [1/50], Iter [10/209] Loss: 44.3375, average_loss: 62.6114
Epoch [1/50], Iter [15/209] Loss: 24.9111, average_loss: 51.8733
Epoch [1/50], Iter [20/209] Loss: 15.8052, average_loss: 43.3599
Epoch [1/50], Iter [25/209] Loss: 10.1352, average_loss: 36.9862
Epoch [1/50], Iter [30/209] Loss: 11.3848, average_loss: 32.8090
Epoch [1/50], Iter [35/209] Loss: 9.6777, average_loss: 29.7235
Epoch [1/50], Iter [40/209] Loss: 9.0094, average_loss: 27.2329
Epoch [1/50], Iter [45/209] Loss: 11.0364, average_loss: 25.2746
Epoch [1/50], Iter [50/209] Loss: 7.2541, average_loss: 23.6024
Epoch [1/50], Iter [55/209] Loss: 6.6169, average_loss: 22.2723
Epoch [1/50], Iter [60/209] Loss: 7.4227, average_loss: 21.1670
Epoch [1/50], Iter [65/209] Loss: 8.2466, average_loss: 20.0769
Epoch [1/50], Iter [70/209] Loss: 6.2321, average_loss: 19.1295
Epoch [1/50], Iter [75/209] Loss: 9.4239, average_loss: 18.3780
Epoch [1/50], Iter [80/209] Loss: 7.4290, average_loss: 17.7499
Epoch [1/50], Iter [85/209] Loss: 7.7592, average_loss: 17.1438
Epoch [1/50], Iter [90/209] Loss: 6.1627, average_loss: 16.6382
Epoch [1/50], Iter [95/209] Loss: 8.0314, average_loss: 16.1569
Epoch [1/50], Iter [100/209] Loss: 5.4507, average_loss: 15.6821
Epoch [1/50], Iter [105/209] Loss: 5.8459, average_loss: 15.2736
Epoch [1/50], Iter [110/209] Loss: 5.8704, average_loss: 14.8524
Epoch [1/50], Iter [115/209] Loss: 5.8754, average_loss: 14.4892
Epoch [1/50], Iter [120/209] Loss: 8.5647, average_loss: 14.1856
Epoch [1/50], Iter [125/209] Loss: 5.5086, average_loss: 13.8970
Epoch [1/50], Iter [130/209] Loss: 8.7413, average_loss: 13.6085
Epoch [1/50], Iter [135/209] Loss: 6.5473, average_loss: 13.3463
Epoch [1/50], Iter [140/209] Loss: 6.7678, average_loss: 13.1446
Epoch [1/50], Iter [145/209] Loss: 6.7875, average_loss: 12.9520
Epoch [1/50], Iter [150/209] Loss: 6.6536, average_loss: 12.7194
Epoch [1/50], Iter [155/209] Loss: 6.4124, average_loss: 12.5249
Epoch [1/50], Iter [160/209] Loss: 6.1811, average_loss: 12.3179
```

MP2_P2-3

March 30, 2020

```
[0]: 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
```

```
[0]: # from google.colab import drive
# drive.mount('/content/drive')
```

0.1 Initialization

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

```
[9]: print(device)
```

cuda:0

```
[0]: # 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.

```
[11]: load_network_path = "best_detector.pth"  
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)
```

Loading saved network from best_detector.pth

```
[0]: 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
```

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

0.2 Reading Pascal Data

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

The train dataset loader also using a variety of data augmentation techniques including random shift, scaling, crop, and flips. Data augmentation is slightly more complicated for detection dataset

since the bounding box annotations must be kept consistent through the transformations.

Since the output of the detector network we train is an $S \times S(B^*5+C)$, 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.

```
[14]: 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

```
[15]: 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

0.3 Train detector

```
[0]: # 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.
#     print("i")

#     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')

```

1 View example predictions

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

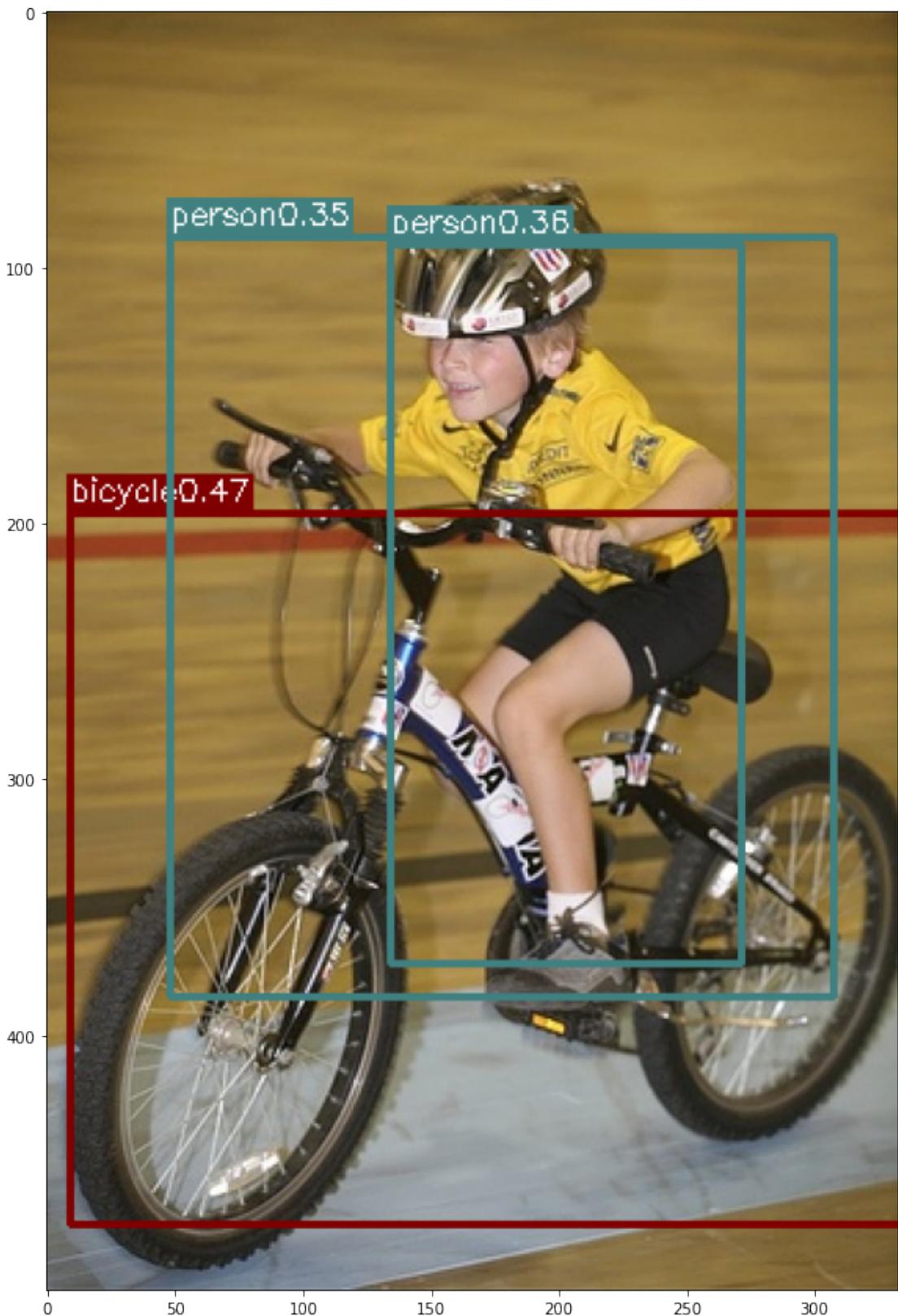
# select random image from test set
image_name = random.choice(test_dataset.fnames)
image = cv2.imread(os.path.join(file_root_test, image_name))
image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)

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...

```
[16]: <matplotlib.image.AxesImage at 0x7fa681f23eb8>
```



```
[17]: # !rmdir VOCdevkit  
# !rm VOCtest_06-Nov-2007.tar
```

```
rmdir: failed to remove 'VOCdevkit': No such file or directory  
rm: cannot remove 'VOCtest_06-Nov-2007.tar': No such file or directory
```

1.1 Evaluate on Test

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

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

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

```
[0]:
```

MP2_P2-3

March 30, 2020

```
[0]: 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
```

```
[0]: # from google.colab import drive
# drive.mount('/content/drive')
```

0.1 Initialization

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

```
[9]: print(device)
```

cuda:0

```
[0]: # 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.

```
[11]: load_network_path = "best_detector.pth"  
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)
```

Loading saved network from best_detector.pth

```
[0]: 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
```

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

0.2 Reading Pascal Data

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

The train dataset loader also using a variety of data augmentation techniques including random shift, scaling, crop, and flips. Data augmentation is slightly more complicated for detection dataset

since the bounding box annotations must be kept consistent through the transformations.

Since the output of the detector network we train is an $S \times S(B^*5+C)$, 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.

```
[14]: 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

```
[15]: 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

0.3 Train detector

```
[0]: # 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.
#     print("i")

#     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')

```

1 View example predictions

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

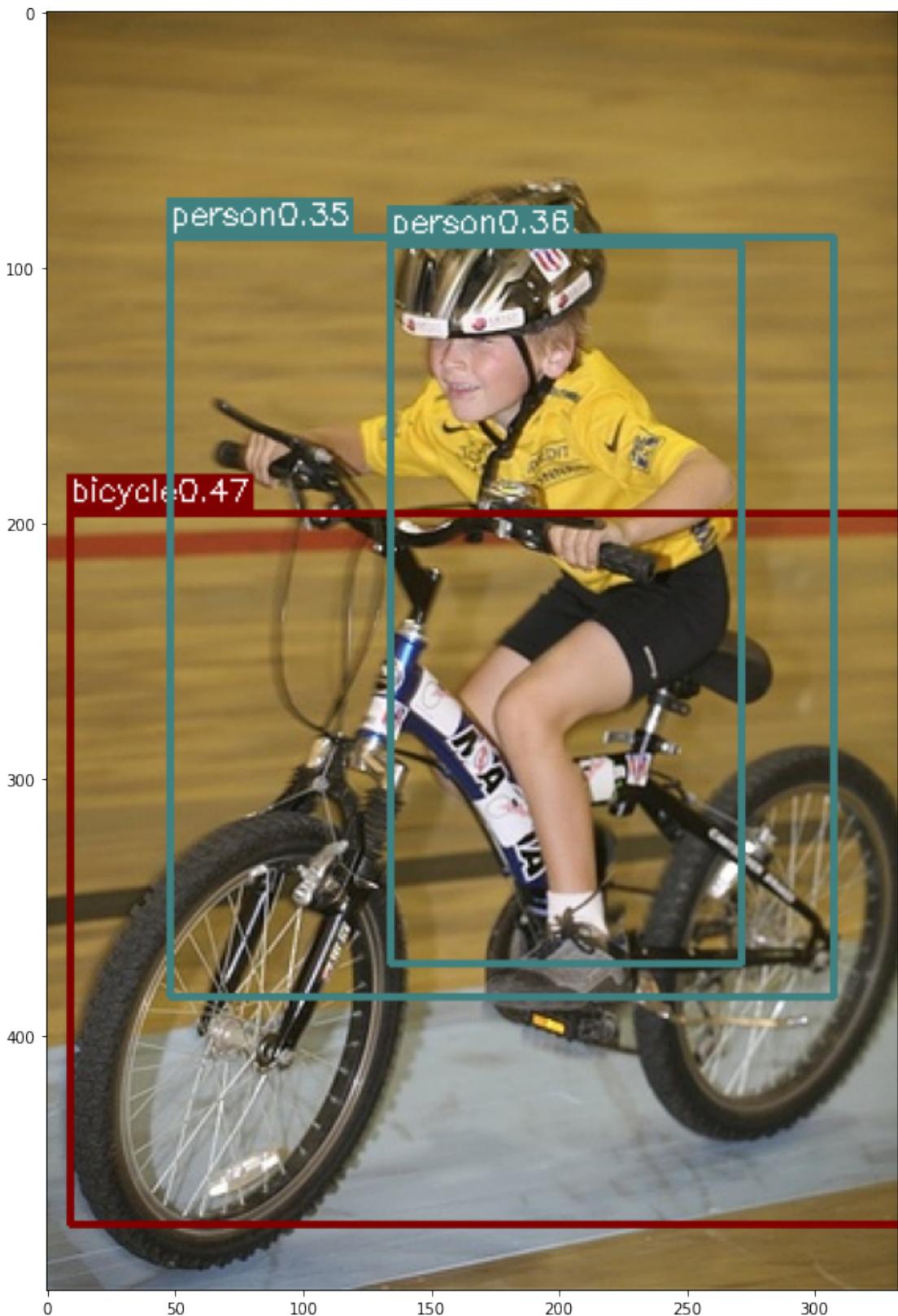
# select random image from test set
image_name = random.choice(test_dataset.fnames)
image = cv2.imread(os.path.join(file_root_test, image_name))
image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)

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...

```
[16]: <matplotlib.image.AxesImage at 0x7fa681f23eb8>
```



```
[17]: # !rmdir VOCdevkit  
# !rm VOCtest_06-Nov-2007.tar
```

```
rmdir: failed to remove 'VOCdevkit': No such file or directory  
rm: cannot remove 'VOCtest_06-Nov-2007.tar': No such file or directory
```

1.1 Evaluate on Test

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

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

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

```
[0]:
```

Epoch [1/50], Iter [165/209] Loss: 5.3245, average_loss: 12.1343
Epoch [1/50], Iter [170/209] Loss: 6.3823, average_loss: 11.9903
Epoch [1/50], Iter [175/209] Loss: 5.5201, average_loss: 11.8107
Epoch [1/50], Iter [180/209] Loss: 6.8564, average_loss: 11.6536
Epoch [1/50], Iter [185/209] Loss: 5.9472, average_loss: 11.4878
Epoch [1/50], Iter [190/209] Loss: 8.8236, average_loss: 11.3597
Epoch [1/50], Iter [195/209] Loss: 5.5406, average_loss: 11.2160
Epoch [1/50], Iter [200/209] Loss: 5.2419, average_loss: 11.0772
Epoch [1/50], Iter [205/209] Loss: 5.6458, average_loss: 10.9566
Updating best test loss: 5.98347

Starting epoch 2 / 50

Learning Rate for this epoch: 0.001

Epoch [2/50], Iter [5/209] Loss: 6.3052, average_loss: 6.7152
Epoch [2/50], Iter [10/209] Loss: 7.6296, average_loss: 6.2041
Epoch [2/50], Iter [15/209] Loss: 5.5551, average_loss: 5.9483
Epoch [2/50], Iter [20/209] Loss: 4.9618, average_loss: 5.9278
Epoch [2/50], Iter [25/209] Loss: 6.0371, average_loss: 5.9408
Epoch [2/50], Iter [30/209] Loss: 8.5306, average_loss: 5.9714
Epoch [2/50], Iter [35/209] Loss: 4.5357, average_loss: 5.9174
Epoch [2/50], Iter [40/209] Loss: 7.5904, average_loss: 6.0012
Epoch [2/50], Iter [45/209] Loss: 6.2147, average_loss: 5.9947
Epoch [2/50], Iter [50/209] Loss: 5.5810, average_loss: 5.9078
Epoch [2/50], Iter [55/209] Loss: 7.3881, average_loss: 5.9337
Epoch [2/50], Iter [60/209] Loss: 4.1955, average_loss: 5.8794
Epoch [2/50], Iter [65/209] Loss: 6.5126, average_loss: 5.8847
Epoch [2/50], Iter [70/209] Loss: 5.4807, average_loss: 5.8623
Epoch [2/50], Iter [75/209] Loss: 5.5140, average_loss: 5.8462
Epoch [2/50], Iter [80/209] Loss: 5.9912, average_loss: 5.8454
Epoch [2/50], Iter [85/209] Loss: 6.7909, average_loss: 5.8551
Epoch [2/50], Iter [90/209] Loss: 5.8773, average_loss: 5.8474
Epoch [2/50], Iter [95/209] Loss: 7.3084, average_loss: 5.8620
Epoch [2/50], Iter [100/209] Loss: 6.1934, average_loss: 5.8411
Epoch [2/50], Iter [105/209] Loss: 6.0227, average_loss: 5.8376
Epoch [2/50], Iter [110/209] Loss: 6.3627, average_loss: 5.8372
Epoch [2/50], Iter [115/209] Loss: 5.0414, average_loss: 5.8156
Epoch [2/50], Iter [120/209] Loss: 5.9797, average_loss: 5.8182
Epoch [2/50], Iter [125/209] Loss: 6.0253, average_loss: 5.8036
Epoch [2/50], Iter [130/209] Loss: 5.1491, average_loss: 5.7769
Epoch [2/50], Iter [135/209] Loss: 6.0885, average_loss: 5.7843
Epoch [2/50], Iter [140/209] Loss: 4.7007, average_loss: 5.7641
Epoch [2/50], Iter [145/209] Loss: 4.9454, average_loss: 5.7700
Epoch [2/50], Iter [150/209] Loss: 4.6005, average_loss: 5.7561
Epoch [2/50], Iter [155/209] Loss: 5.4243, average_loss: 5.7514
Epoch [2/50], Iter [160/209] Loss: 4.2194, average_loss: 5.7437
Epoch [2/50], Iter [165/209] Loss: 4.5435, average_loss: 5.7131
Epoch [2/50], Iter [170/209] Loss: 5.4181, average_loss: 5.7129
Epoch [2/50], Iter [175/209] Loss: 5.4562, average_loss: 5.7083
Epoch [2/50], Iter [180/209] Loss: 4.3098, average_loss: 5.7087
Epoch [2/50], Iter [185/209] Loss: 5.4097, average_loss: 5.7076
Epoch [2/50], Iter [190/209] Loss: 5.1185, average_loss: 5.6933
Epoch [2/50], Iter [195/209] Loss: 4.5494, average_loss: 5.6904
Epoch [2/50], Iter [200/209] Loss: 5.7090, average_loss: 5.7056

MP2_P2-3

March 30, 2020

```
[0]: 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
```

```
[0]: # from google.colab import drive
# drive.mount('/content/drive')
```

0.1 Initialization

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

```
[9]: print(device)
```

cuda:0

```
[0]: # 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.

```
[11]: load_network_path = "best_detector.pth"  
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)
```

Loading saved network from best_detector.pth

```
[0]: 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
```

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

0.2 Reading Pascal Data

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

The train dataset loader also using a variety of data augmentation techniques including random shift, scaling, crop, and flips. Data augmentation is slightly more complicated for detection dataset

since the bounding box annotations must be kept consistent through the transformations.

Since the output of the detector network we train is an $S \times S(B^*5+C)$, 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.

```
[14]: 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

```
[15]: 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

0.3 Train detector

```
[0]: # 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.
#     print("i")

#     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')

```

1 View example predictions

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

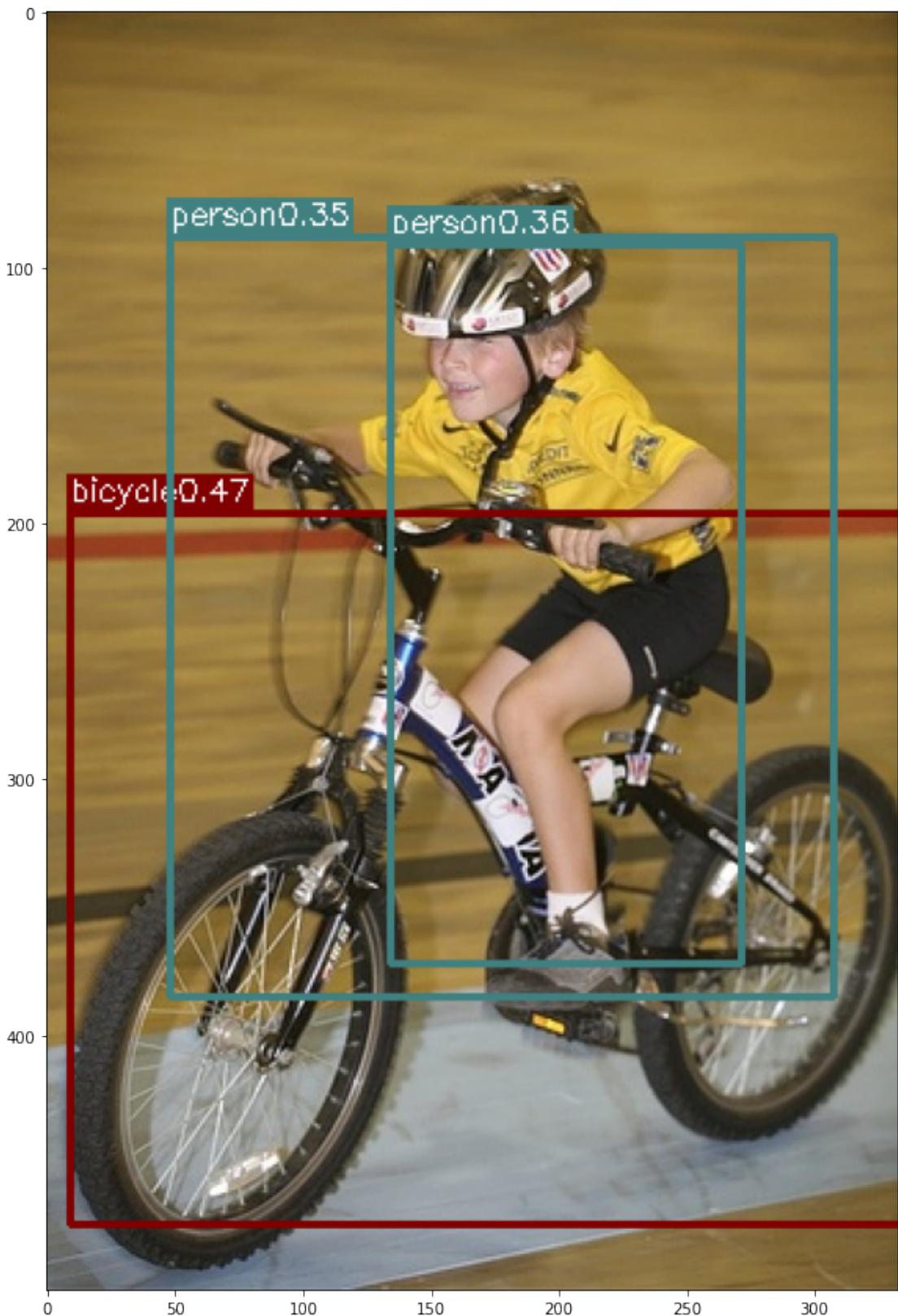
# select random image from test set
image_name = random.choice(test_dataset.fnames)
image = cv2.imread(os.path.join(file_root_test, image_name))
image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)

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...

```
[16]: <matplotlib.image.AxesImage at 0x7fa681f23eb8>
```



```
[17]: # !rmdir VOCdevkit  
# !rm VOCtest_06-Nov-2007.tar
```

```
rmdir: failed to remove 'VOCdevkit': No such file or directory  
rm: cannot remove 'VOCtest_06-Nov-2007.tar': No such file or directory
```

1.1 Evaluate on Test

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

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

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

```
[0]:
```

Epoch [2/50], Iter [205/209] Loss: 5.2568, average_loss: 5.6967
Updating best test loss: 5.43809

Starting epoch 3 / 50

Learning Rate for this epoch: 0.001

Epoch [3/50], Iter [5/209] Loss: 6.4025, average_loss: 5.6884
Epoch [3/50], Iter [10/209] Loss: 5.3757, average_loss: 5.6190
Epoch [3/50], Iter [15/209] Loss: 4.9492, average_loss: 5.2300
Epoch [3/50], Iter [20/209] Loss: 5.0341, average_loss: 5.2154
Epoch [3/50], Iter [25/209] Loss: 5.9862, average_loss: 5.3076
Epoch [3/50], Iter [30/209] Loss: 6.3620, average_loss: 5.3373
Epoch [3/50], Iter [35/209] Loss: 5.9831, average_loss: 5.3245
Epoch [3/50], Iter [40/209] Loss: 5.2239, average_loss: 5.3942
Epoch [3/50], Iter [45/209] Loss: 4.9064, average_loss: 5.3319
Epoch [3/50], Iter [50/209] Loss: 5.7398, average_loss: 5.3181
Epoch [3/50], Iter [55/209] Loss: 5.4480, average_loss: 5.3169
Epoch [3/50], Iter [60/209] Loss: 4.5123, average_loss: 5.2783
Epoch [3/50], Iter [65/209] Loss: 4.1824, average_loss: 5.2623
Epoch [3/50], Iter [70/209] Loss: 5.0689, average_loss: 5.2705
Epoch [3/50], Iter [75/209] Loss: 5.3328, average_loss: 5.2568
Epoch [3/50], Iter [80/209] Loss: 6.6308, average_loss: 5.3395
Epoch [3/50], Iter [85/209] Loss: 4.8163, average_loss: 5.3234
Epoch [3/50], Iter [90/209] Loss: 4.6629, average_loss: 5.3365
Epoch [3/50], Iter [95/209] Loss: 4.5669, average_loss: 5.3231
Epoch [3/50], Iter [100/209] Loss: 5.7914, average_loss: 5.3080
Epoch [3/50], Iter [105/209] Loss: 4.9918, average_loss: 5.3124
Epoch [3/50], Iter [110/209] Loss: 4.7263, average_loss: 5.3002
Epoch [3/50], Iter [115/209] Loss: 5.3043, average_loss: 5.2959
Epoch [3/50], Iter [120/209] Loss: 4.1370, average_loss: 5.2832
Epoch [3/50], Iter [125/209] Loss: 5.2974, average_loss: 5.2779
Epoch [3/50], Iter [130/209] Loss: 5.6775, average_loss: 5.2746
Epoch [3/50], Iter [135/209] Loss: 4.4840, average_loss: 5.2680
Epoch [3/50], Iter [140/209] Loss: 5.1921, average_loss: 5.2558
Epoch [3/50], Iter [145/209] Loss: 5.0880, average_loss: 5.2367
Epoch [3/50], Iter [150/209] Loss: 4.2985, average_loss: 5.2234
Epoch [3/50], Iter [155/209] Loss: 3.6069, average_loss: 5.2144
Epoch [3/50], Iter [160/209] Loss: 4.7670, average_loss: 5.1985
Epoch [3/50], Iter [165/209] Loss: 4.1101, average_loss: 5.2012
Epoch [3/50], Iter [170/209] Loss: 5.9430, average_loss: 5.1959
Epoch [3/50], Iter [175/209] Loss: 4.1853, average_loss: 5.1939
Epoch [3/50], Iter [180/209] Loss: 5.5167, average_loss: 5.2027
Epoch [3/50], Iter [185/209] Loss: 5.9659, average_loss: 5.2065
Epoch [3/50], Iter [190/209] Loss: 5.5232, average_loss: 5.2014
Epoch [3/50], Iter [195/209] Loss: 5.7552, average_loss: 5.2006
Epoch [3/50], Iter [200/209] Loss: 4.6445, average_loss: 5.1944
Epoch [3/50], Iter [205/209] Loss: 4.9723, average_loss: 5.1866
Updating best test loss: 5.18997

Starting epoch 4 / 50

Learning Rate for this epoch: 0.001

Epoch [4/50], Iter [5/209] Loss: 4.4102, average_loss: 4.5622
Epoch [4/50], Iter [10/209] Loss: 6.4812, average_loss: 4.8595

Epoch [4/50], Iter [15/209] Loss: 5.2445, average_loss: 4.8378
Epoch [4/50], Iter [20/209] Loss: 5.1538, average_loss: 4.8143
Epoch [4/50], Iter [25/209] Loss: 5.0795, average_loss: 4.8070
Epoch [4/50], Iter [30/209] Loss: 3.9242, average_loss: 4.7762
Epoch [4/50], Iter [35/209] Loss: 4.3037, average_loss: 4.8247
Epoch [4/50], Iter [40/209] Loss: 4.3284, average_loss: 4.7997
Epoch [4/50], Iter [45/209] Loss: 4.3391, average_loss: 4.8098
Epoch [4/50], Iter [50/209] Loss: 3.8220, average_loss: 4.7959
Epoch [4/50], Iter [55/209] Loss: 4.3369, average_loss: 4.7999
Epoch [4/50], Iter [60/209] Loss: 5.2191, average_loss: 4.7855
Epoch [4/50], Iter [65/209] Loss: 3.4048, average_loss: 4.7735
Epoch [4/50], Iter [70/209] Loss: 8.4940, average_loss: 4.8730
Epoch [4/50], Iter [75/209] Loss: 4.2000, average_loss: 4.8638
Epoch [4/50], Iter [80/209] Loss: 5.2955, average_loss: 4.8549
Epoch [4/50], Iter [85/209] Loss: 5.1259, average_loss: 4.8561
Epoch [4/50], Iter [90/209] Loss: 5.2663, average_loss: 4.8738
Epoch [4/50], Iter [95/209] Loss: 4.6420, average_loss: 4.8706
Epoch [4/50], Iter [100/209] Loss: 4.5200, average_loss: 4.8375
Epoch [4/50], Iter [105/209] Loss: 4.7462, average_loss: 4.8487
Epoch [4/50], Iter [110/209] Loss: 6.6480, average_loss: 4.8664
Epoch [4/50], Iter [115/209] Loss: 4.9502, average_loss: 4.8590
Epoch [4/50], Iter [120/209] Loss: 3.9032, average_loss: 4.8595
Epoch [4/50], Iter [125/209] Loss: 4.7689, average_loss: 4.8778
Epoch [4/50], Iter [130/209] Loss: 5.4185, average_loss: 4.8864
Epoch [4/50], Iter [135/209] Loss: 4.5608, average_loss: 4.8695
Epoch [4/50], Iter [140/209] Loss: 5.1421, average_loss: 4.8794
Epoch [4/50], Iter [145/209] Loss: 6.2481, average_loss: 4.8667
Epoch [4/50], Iter [150/209] Loss: 4.2092, average_loss: 4.8632
Epoch [4/50], Iter [155/209] Loss: 5.2434, average_loss: 4.8631
Epoch [4/50], Iter [160/209] Loss: 5.0977, average_loss: 4.8468
Epoch [4/50], Iter [165/209] Loss: 5.0512, average_loss: 4.8408
Epoch [4/50], Iter [170/209] Loss: 3.9531, average_loss: 4.8306
Epoch [4/50], Iter [175/209] Loss: 4.1557, average_loss: 4.8399
Epoch [4/50], Iter [180/209] Loss: 5.1020, average_loss: 4.8444
Epoch [4/50], Iter [185/209] Loss: 4.4546, average_loss: 4.8305
Epoch [4/50], Iter [190/209] Loss: 4.5837, average_loss: 4.8229
Epoch [4/50], Iter [195/209] Loss: 5.2478, average_loss: 4.8226
Epoch [4/50], Iter [200/209] Loss: 5.5048, average_loss: 4.8164
Epoch [4/50], Iter [205/209] Loss: 3.6468, average_loss: 4.8068
Updating best test loss: 4.84918

Starting epoch 5 / 50

Learning Rate for this epoch: 0.001

Epoch [5/50], Iter [5/209] Loss: 4.7013, average_loss: 4.3272
Epoch [5/50], Iter [10/209] Loss: 3.9175, average_loss: 4.3117
Epoch [5/50], Iter [15/209] Loss: 4.5514, average_loss: 4.3771
Epoch [5/50], Iter [20/209] Loss: 3.9162, average_loss: 4.4506
Epoch [5/50], Iter [25/209] Loss: 4.1954, average_loss: 4.3882
Epoch [5/50], Iter [30/209] Loss: 3.9153, average_loss: 4.4339
Epoch [5/50], Iter [35/209] Loss: 3.9736, average_loss: 4.4352
Epoch [5/50], Iter [40/209] Loss: 3.5608, average_loss: 4.4593
Epoch [5/50], Iter [45/209] Loss: 4.2625, average_loss: 4.4973
Epoch [5/50], Iter [50/209] Loss: 5.4487, average_loss: 4.4943

Epoch [5/50], Iter [55/209] Loss: 4.0980, average_loss: 4.4910
Epoch [5/50], Iter [60/209] Loss: 4.3959, average_loss: 4.4503
Epoch [5/50], Iter [65/209] Loss: 4.3639, average_loss: 4.4332
Epoch [5/50], Iter [70/209] Loss: 4.4559, average_loss: 4.4738
Epoch [5/50], Iter [75/209] Loss: 3.8859, average_loss: 4.4334
Epoch [5/50], Iter [80/209] Loss: 4.8998, average_loss: 4.4587
Epoch [5/50], Iter [85/209] Loss: 3.7936, average_loss: 4.4386
Epoch [5/50], Iter [90/209] Loss: 3.8360, average_loss: 4.4259
Epoch [5/50], Iter [95/209] Loss: 4.0782, average_loss: 4.4311
Epoch [5/50], Iter [100/209] Loss: 3.3771, average_loss: 4.4111
Epoch [5/50], Iter [105/209] Loss: 5.6158, average_loss: 4.4150
Epoch [5/50], Iter [110/209] Loss: 3.6636, average_loss: 4.4385
Epoch [5/50], Iter [115/209] Loss: 4.6080, average_loss: 4.4369
Epoch [5/50], Iter [120/209] Loss: 4.0216, average_loss: 4.4198
Epoch [5/50], Iter [125/209] Loss: 4.5501, average_loss: 4.4003
Epoch [5/50], Iter [130/209] Loss: 3.5867, average_loss: 4.3954
Epoch [5/50], Iter [135/209] Loss: 4.5756, average_loss: 4.4054
Epoch [5/50], Iter [140/209] Loss: 5.7831, average_loss: 4.4195
Epoch [5/50], Iter [145/209] Loss: 5.2654, average_loss: 4.4375
Epoch [5/50], Iter [150/209] Loss: 4.1743, average_loss: 4.4566
Epoch [5/50], Iter [155/209] Loss: 4.0706, average_loss: 4.4428
Epoch [5/50], Iter [160/209] Loss: 4.4991, average_loss: 4.4369
Epoch [5/50], Iter [165/209] Loss: 5.4118, average_loss: 4.4551
Epoch [5/50], Iter [170/209] Loss: 4.9744, average_loss: 4.4551
Epoch [5/50], Iter [175/209] Loss: 5.5762, average_loss: 4.4724
Epoch [5/50], Iter [180/209] Loss: 5.0354, average_loss: 4.5023
Epoch [5/50], Iter [185/209] Loss: 4.3536, average_loss: 4.5027
Epoch [5/50], Iter [190/209] Loss: 4.5772, average_loss: 4.5103
Epoch [5/50], Iter [195/209] Loss: 3.1988, average_loss: 4.4879
Epoch [5/50], Iter [200/209] Loss: 3.9523, average_loss: 4.4874
Epoch [5/50], Iter [205/209] Loss: 3.9763, average_loss: 4.4731
Updating best test loss: 4.52716

Starting epoch 6 / 50

Learning Rate for this epoch: 0.001

Epoch [6/50], Iter [5/209] Loss: 5.4945, average_loss: 4.9836
Epoch [6/50], Iter [10/209] Loss: 4.2077, average_loss: 5.0588
Epoch [6/50], Iter [15/209] Loss: 3.8606, average_loss: 4.6726
Epoch [6/50], Iter [20/209] Loss: 4.5956, average_loss: 4.6774
Epoch [6/50], Iter [25/209] Loss: 4.7153, average_loss: 4.6315
Epoch [6/50], Iter [30/209] Loss: 4.1448, average_loss: 4.5167
Epoch [6/50], Iter [35/209] Loss: 4.7924, average_loss: 4.4945
Epoch [6/50], Iter [40/209] Loss: 4.5983, average_loss: 4.4150
Epoch [6/50], Iter [45/209] Loss: 4.7084, average_loss: 4.3939
Epoch [6/50], Iter [50/209] Loss: 3.4544, average_loss: 4.3982
Epoch [6/50], Iter [55/209] Loss: 3.0618, average_loss: 4.3347
Epoch [6/50], Iter [60/209] Loss: 3.4658, average_loss: 4.3055
Epoch [6/50], Iter [65/209] Loss: 4.3029, average_loss: 4.3259
Epoch [6/50], Iter [70/209] Loss: 4.6507, average_loss: 4.3217
Epoch [6/50], Iter [75/209] Loss: 3.4686, average_loss: 4.3066
Epoch [6/50], Iter [80/209] Loss: 4.1904, average_loss: 4.2894
Epoch [6/50], Iter [85/209] Loss: 3.6176, average_loss: 4.2612
Epoch [6/50], Iter [90/209] Loss: 4.9458, average_loss: 4.2614

Epoch [6/50], Iter [95/209] Loss: 3.6226, average_loss: 4.2669
Epoch [6/50], Iter [100/209] Loss: 3.9876, average_loss: 4.2722
Epoch [6/50], Iter [105/209] Loss: 3.9531, average_loss: 4.2492
Epoch [6/50], Iter [110/209] Loss: 5.1425, average_loss: 4.2550
Epoch [6/50], Iter [115/209] Loss: 3.5101, average_loss: 4.2389
Epoch [6/50], Iter [120/209] Loss: 4.2411, average_loss: 4.2204
Epoch [6/50], Iter [125/209] Loss: 2.8572, average_loss: 4.2300
Epoch [6/50], Iter [130/209] Loss: 3.9224, average_loss: 4.2406
Epoch [6/50], Iter [135/209] Loss: 4.4888, average_loss: 4.2221
Epoch [6/50], Iter [140/209] Loss: 4.9217, average_loss: 4.2257
Epoch [6/50], Iter [145/209] Loss: 4.4163, average_loss: 4.2335
Epoch [6/50], Iter [150/209] Loss: 4.5547, average_loss: 4.2323
Epoch [6/50], Iter [155/209] Loss: 5.5432, average_loss: 4.2345
Epoch [6/50], Iter [160/209] Loss: 4.5502, average_loss: 4.2327
Epoch [6/50], Iter [165/209] Loss: 4.1402, average_loss: 4.2306
Epoch [6/50], Iter [170/209] Loss: 2.6548, average_loss: 4.2404
Epoch [6/50], Iter [175/209] Loss: 5.1323, average_loss: 4.2491
Epoch [6/50], Iter [180/209] Loss: 2.8421, average_loss: 4.2368
Epoch [6/50], Iter [185/209] Loss: 3.0422, average_loss: 4.2272
Epoch [6/50], Iter [190/209] Loss: 2.9975, average_loss: 4.2189
Epoch [6/50], Iter [195/209] Loss: 5.3416, average_loss: 4.2253
Epoch [6/50], Iter [200/209] Loss: 5.6619, average_loss: 4.2202
Epoch [6/50], Iter [205/209] Loss: 3.8271, average_loss: 4.2206
Updating best test loss: 4.34128

Starting epoch 7 / 50

Learning Rate for this epoch: 0.001

Epoch [7/50], Iter [5/209] Loss: 4.4561, average_loss: 4.0863
Epoch [7/50], Iter [10/209] Loss: 3.5869, average_loss: 4.0185
Epoch [7/50], Iter [15/209] Loss: 4.1338, average_loss: 3.9575
Epoch [7/50], Iter [20/209] Loss: 4.4054, average_loss: 3.8877
Epoch [7/50], Iter [25/209] Loss: 3.4470, average_loss: 3.9387
Epoch [7/50], Iter [30/209] Loss: 4.1238, average_loss: 3.9291
Epoch [7/50], Iter [35/209] Loss: 4.6714, average_loss: 3.9916
Epoch [7/50], Iter [40/209] Loss: 4.4817, average_loss: 4.0035
Epoch [7/50], Iter [45/209] Loss: 3.8750, average_loss: 3.9867
Epoch [7/50], Iter [50/209] Loss: 3.2809, average_loss: 4.0161
Epoch [7/50], Iter [55/209] Loss: 3.2522, average_loss: 3.9878
Epoch [7/50], Iter [60/209] Loss: 5.3099, average_loss: 4.0013
Epoch [7/50], Iter [65/209] Loss: 4.6731, average_loss: 4.0091
Epoch [7/50], Iter [70/209] Loss: 6.5167, average_loss: 4.0432
Epoch [7/50], Iter [75/209] Loss: 3.5070, average_loss: 4.0418
Epoch [7/50], Iter [80/209] Loss: 5.8695, average_loss: 4.0549
Epoch [7/50], Iter [85/209] Loss: 3.4555, average_loss: 4.0515
Epoch [7/50], Iter [90/209] Loss: 3.9826, average_loss: 4.0351
Epoch [7/50], Iter [95/209] Loss: 3.9350, average_loss: 4.0469
Epoch [7/50], Iter [100/209] Loss: 5.0305, average_loss: 4.0766
Epoch [7/50], Iter [105/209] Loss: 5.0873, average_loss: 4.0915
Epoch [7/50], Iter [110/209] Loss: 3.9589, average_loss: 4.0733
Epoch [7/50], Iter [115/209] Loss: 3.2767, average_loss: 4.0749
Epoch [7/50], Iter [120/209] Loss: 4.3749, average_loss: 4.0679
Epoch [7/50], Iter [125/209] Loss: 5.0794, average_loss: 4.0716
Epoch [7/50], Iter [130/209] Loss: 4.0448, average_loss: 4.0807

Epoch [7/50], Iter [135/209] Loss: 3.8499, average_loss: 4.0792
Epoch [7/50], Iter [140/209] Loss: 3.6383, average_loss: 4.0731
Epoch [7/50], Iter [145/209] Loss: 5.1877, average_loss: 4.0784
Epoch [7/50], Iter [150/209] Loss: 3.8274, average_loss: 4.0744
Epoch [7/50], Iter [155/209] Loss: 4.4478, average_loss: 4.0711
Epoch [7/50], Iter [160/209] Loss: 3.8898, average_loss: 4.0690
Epoch [7/50], Iter [165/209] Loss: 3.2475, average_loss: 4.0684
Epoch [7/50], Iter [170/209] Loss: 4.3867, average_loss: 4.0646
Epoch [7/50], Iter [175/209] Loss: 5.0229, average_loss: 4.0639
Epoch [7/50], Iter [180/209] Loss: 5.1454, average_loss: 4.0615
Epoch [7/50], Iter [185/209] Loss: 4.6575, average_loss: 4.0630
Epoch [7/50], Iter [190/209] Loss: 3.4091, average_loss: 4.0525
Epoch [7/50], Iter [195/209] Loss: 3.6600, average_loss: 4.0573
Epoch [7/50], Iter [200/209] Loss: 6.5462, average_loss: 4.0639
Epoch [7/50], Iter [205/209] Loss: 3.5426, average_loss: 4.0636
Updating best test loss: 4.21879

Starting epoch 8 / 50

Learning Rate for this epoch: 0.001

Epoch [8/50], Iter [5/209] Loss: 3.8637, average_loss: 4.0507
Epoch [8/50], Iter [10/209] Loss: 3.5448, average_loss: 4.0127
Epoch [8/50], Iter [15/209] Loss: 3.3692, average_loss: 3.8196
Epoch [8/50], Iter [20/209] Loss: 3.2704, average_loss: 3.8381
Epoch [8/50], Iter [25/209] Loss: 3.5758, average_loss: 3.8366
Epoch [8/50], Iter [30/209] Loss: 3.7387, average_loss: 3.8372
Epoch [8/50], Iter [35/209] Loss: 3.5672, average_loss: 3.8165
Epoch [8/50], Iter [40/209] Loss: 2.9874, average_loss: 3.7846
Epoch [8/50], Iter [45/209] Loss: 3.9338, average_loss: 3.8540
Epoch [8/50], Iter [50/209] Loss: 3.4375, average_loss: 3.8851
Epoch [8/50], Iter [55/209] Loss: 3.9786, average_loss: 3.8877
Epoch [8/50], Iter [60/209] Loss: 3.1096, average_loss: 3.8618
Epoch [8/50], Iter [65/209] Loss: 3.7974, average_loss: 3.8640
Epoch [8/50], Iter [70/209] Loss: 3.8520, average_loss: 3.8655
Epoch [8/50], Iter [75/209] Loss: 4.4325, average_loss: 3.8603
Epoch [8/50], Iter [80/209] Loss: 4.0717, average_loss: 3.8795
Epoch [8/50], Iter [85/209] Loss: 3.2061, average_loss: 3.8543
Epoch [8/50], Iter [90/209] Loss: 4.7716, average_loss: 3.9209
Epoch [8/50], Iter [95/209] Loss: 3.6668, average_loss: 3.9180
Epoch [8/50], Iter [100/209] Loss: 4.2576, average_loss: 3.9235
Epoch [8/50], Iter [105/209] Loss: 3.9511, average_loss: 3.9114
Epoch [8/50], Iter [110/209] Loss: 4.3481, average_loss: 3.9113
Epoch [8/50], Iter [115/209] Loss: 4.5301, average_loss: 3.9305
Epoch [8/50], Iter [120/209] Loss: 3.8635, average_loss: 3.9222
Epoch [8/50], Iter [125/209] Loss: 4.2186, average_loss: 3.9266
Epoch [8/50], Iter [130/209] Loss: 3.2656, average_loss: 3.9175
Epoch [8/50], Iter [135/209] Loss: 3.4143, average_loss: 3.9074
Epoch [8/50], Iter [140/209] Loss: 3.5435, average_loss: 3.8863
Epoch [8/50], Iter [145/209] Loss: 3.4475, average_loss: 3.8789
Epoch [8/50], Iter [150/209] Loss: 3.8779, average_loss: 3.8731
Epoch [8/50], Iter [155/209] Loss: 4.0907, average_loss: 3.8838
Epoch [8/50], Iter [160/209] Loss: 3.6815, average_loss: 3.8823
Epoch [8/50], Iter [165/209] Loss: 2.5136, average_loss: 3.8719
Epoch [8/50], Iter [170/209] Loss: 3.2118, average_loss: 3.8690

Epoch [8/50], Iter [175/209] Loss: 4.2414, average_loss: 3.8628
Epoch [8/50], Iter [180/209] Loss: 2.7612, average_loss: 3.8718
Epoch [8/50], Iter [185/209] Loss: 3.8527, average_loss: 3.8723
Epoch [8/50], Iter [190/209] Loss: 3.7286, average_loss: 3.8585
Epoch [8/50], Iter [195/209] Loss: 2.9305, average_loss: 3.8373
Epoch [8/50], Iter [200/209] Loss: 4.0610, average_loss: 3.8366
Epoch [8/50], Iter [205/209] Loss: 4.3327, average_loss: 3.8469
Updating best test loss: 4.15904

Starting epoch 9 / 50

Learning Rate for this epoch: 0.001

Epoch [9/50], Iter [5/209] Loss: 3.2376, average_loss: 3.5561
Epoch [9/50], Iter [10/209] Loss: 2.9998, average_loss: 3.5604
Epoch [9/50], Iter [15/209] Loss: 5.2370, average_loss: 3.8176
Epoch [9/50], Iter [20/209] Loss: 3.3171, average_loss: 3.8304
Epoch [9/50], Iter [25/209] Loss: 2.9820, average_loss: 3.7406
Epoch [9/50], Iter [30/209] Loss: 3.5953, average_loss: 3.7593
Epoch [9/50], Iter [35/209] Loss: 4.3275, average_loss: 3.8073
Epoch [9/50], Iter [40/209] Loss: 3.6425, average_loss: 3.8125
Epoch [9/50], Iter [45/209] Loss: 3.4972, average_loss: 3.8520
Epoch [9/50], Iter [50/209] Loss: 3.2589, average_loss: 3.8674
Epoch [9/50], Iter [55/209] Loss: 3.7236, average_loss: 3.8191
Epoch [9/50], Iter [60/209] Loss: 3.3215, average_loss: 3.8037
Epoch [9/50], Iter [65/209] Loss: 3.1770, average_loss: 3.7959
Epoch [9/50], Iter [70/209] Loss: 3.8289, average_loss: 3.7988
Epoch [9/50], Iter [75/209] Loss: 3.7340, average_loss: 3.7782
Epoch [9/50], Iter [80/209] Loss: 3.7565, average_loss: 3.7676
Epoch [9/50], Iter [85/209] Loss: 3.6321, average_loss: 3.7635
Epoch [9/50], Iter [90/209] Loss: 4.5791, average_loss: 3.8165
Epoch [9/50], Iter [95/209] Loss: 2.5623, average_loss: 3.7859
Epoch [9/50], Iter [100/209] Loss: 3.0947, average_loss: 3.7940
Epoch [9/50], Iter [105/209] Loss: 2.8004, average_loss: 3.7591
Epoch [9/50], Iter [110/209] Loss: 2.7871, average_loss: 3.7622
Epoch [9/50], Iter [115/209] Loss: 3.9493, average_loss: 3.7727
Epoch [9/50], Iter [120/209] Loss: 3.7750, average_loss: 3.7679
Epoch [9/50], Iter [125/209] Loss: 3.6590, average_loss: 3.7680
Epoch [9/50], Iter [130/209] Loss: 3.4680, average_loss: 3.7552
Epoch [9/50], Iter [135/209] Loss: 4.2485, average_loss: 3.7478
Epoch [9/50], Iter [140/209] Loss: 4.1699, average_loss: 3.7448
Epoch [9/50], Iter [145/209] Loss: 3.3143, average_loss: 3.7413
Epoch [9/50], Iter [150/209] Loss: 3.1154, average_loss: 3.7526
Epoch [9/50], Iter [155/209] Loss: 4.0015, average_loss: 3.7553
Epoch [9/50], Iter [160/209] Loss: 2.7427, average_loss: 3.7582
Epoch [9/50], Iter [165/209] Loss: 5.3845, average_loss: 3.7553
Epoch [9/50], Iter [170/209] Loss: 3.2360, average_loss: 3.7394
Epoch [9/50], Iter [175/209] Loss: 3.5071, average_loss: 3.7434
Epoch [9/50], Iter [180/209] Loss: 4.1457, average_loss: 3.7512
Epoch [9/50], Iter [185/209] Loss: 5.1994, average_loss: 3.7540
Epoch [9/50], Iter [190/209] Loss: 3.8876, average_loss: 3.7544
Epoch [9/50], Iter [195/209] Loss: 2.7023, average_loss: 3.7462
Epoch [9/50], Iter [200/209] Loss: 4.2032, average_loss: 3.7557
Epoch [9/50], Iter [205/209] Loss: 4.0449, average_loss: 3.7599
Updating best test loss: 4.04351

Starting epoch 10 / 50

Learning Rate for this epoch: 0.001

Epoch [10/50], Iter [5/209] Loss: 3.0352, average_loss: 3.3614
Epoch [10/50], Iter [10/209] Loss: 2.7097, average_loss: 3.3892
Epoch [10/50], Iter [15/209] Loss: 3.4869, average_loss: 3.6103
Epoch [10/50], Iter [20/209] Loss: 3.5094, average_loss: 3.6896
Epoch [10/50], Iter [25/209] Loss: 4.5489, average_loss: 3.7236
Epoch [10/50], Iter [30/209] Loss: 3.2731, average_loss: 3.6764
Epoch [10/50], Iter [35/209] Loss: 3.3570, average_loss: 3.6863
Epoch [10/50], Iter [40/209] Loss: 3.6412, average_loss: 3.6625
Epoch [10/50], Iter [45/209] Loss: 3.9120, average_loss: 3.6461
Epoch [10/50], Iter [50/209] Loss: 3.7178, average_loss: 3.6222
Epoch [10/50], Iter [55/209] Loss: 3.1872, average_loss: 3.5902
Epoch [10/50], Iter [60/209] Loss: 3.4706, average_loss: 3.6051
Epoch [10/50], Iter [65/209] Loss: 4.0184, average_loss: 3.6298
Epoch [10/50], Iter [70/209] Loss: 3.8341, average_loss: 3.6243
Epoch [10/50], Iter [75/209] Loss: 3.7293, average_loss: 3.6193
Epoch [10/50], Iter [80/209] Loss: 4.4773, average_loss: 3.6490
Epoch [10/50], Iter [85/209] Loss: 2.9581, average_loss: 3.6247
Epoch [10/50], Iter [90/209] Loss: 3.6385, average_loss: 3.6562
Epoch [10/50], Iter [95/209] Loss: 2.9485, average_loss: 3.6484
Epoch [10/50], Iter [100/209] Loss: 5.2807, average_loss: 3.6705
Epoch [10/50], Iter [105/209] Loss: 3.6224, average_loss: 3.6588
Epoch [10/50], Iter [110/209] Loss: 4.2609, average_loss: 3.6651
Epoch [10/50], Iter [115/209] Loss: 4.5807, average_loss: 3.6584
Epoch [10/50], Iter [120/209] Loss: 4.3791, average_loss: 3.6671
Epoch [10/50], Iter [125/209] Loss: 3.7488, average_loss: 3.6639
Epoch [10/50], Iter [130/209] Loss: 3.0173, average_loss: 3.6478
Epoch [10/50], Iter [135/209] Loss: 4.3079, average_loss: 3.6371
Epoch [10/50], Iter [140/209] Loss: 3.9952, average_loss: 3.6461
Epoch [10/50], Iter [145/209] Loss: 3.2829, average_loss: 3.6603
Epoch [10/50], Iter [150/209] Loss: 2.3676, average_loss: 3.6658
Epoch [10/50], Iter [155/209] Loss: 3.6696, average_loss: 3.6718
Epoch [10/50], Iter [160/209] Loss: 3.5007, average_loss: 3.6652
Epoch [10/50], Iter [165/209] Loss: 3.0374, average_loss: 3.6651
Epoch [10/50], Iter [170/209] Loss: 4.1290, average_loss: 3.6625
Epoch [10/50], Iter [175/209] Loss: 3.1489, average_loss: 3.6624
Epoch [10/50], Iter [180/209] Loss: 4.3139, average_loss: 3.6597
Epoch [10/50], Iter [185/209] Loss: 3.3621, average_loss: 3.6510
Epoch [10/50], Iter [190/209] Loss: 4.5960, average_loss: 3.6481
Epoch [10/50], Iter [195/209] Loss: 3.7480, average_loss: 3.6490
Epoch [10/50], Iter [200/209] Loss: 3.5829, average_loss: 3.6449
Epoch [10/50], Iter [205/209] Loss: 3.8990, average_loss: 3.6455
Updating best test loss: 3.99523

Starting epoch 11 / 50

Learning Rate for this epoch: 0.001

Epoch [11/50], Iter [5/209] Loss: 3.3942, average_loss: 3.7524
Epoch [11/50], Iter [10/209] Loss: 3.2413, average_loss: 3.7314
Epoch [11/50], Iter [15/209] Loss: 3.8182, average_loss: 3.6908
Epoch [11/50], Iter [20/209] Loss: 4.0770, average_loss: 3.6344

Epoch [11/50], Iter [25/209] Loss: 2.7420, average_loss: 3.7009
Epoch [11/50], Iter [30/209] Loss: 3.4083, average_loss: 3.6624
Epoch [11/50], Iter [35/209] Loss: 3.0529, average_loss: 3.6694
Epoch [11/50], Iter [40/209] Loss: 3.7186, average_loss: 3.6765
Epoch [11/50], Iter [45/209] Loss: 2.9845, average_loss: 3.6042
Epoch [11/50], Iter [50/209] Loss: 2.8005, average_loss: 3.6253
Epoch [11/50], Iter [55/209] Loss: 3.2541, average_loss: 3.6343
Epoch [11/50], Iter [60/209] Loss: 3.6211, average_loss: 3.6196
Epoch [11/50], Iter [65/209] Loss: 3.5397, average_loss: 3.6042
Epoch [11/50], Iter [70/209] Loss: 3.3742, average_loss: 3.5921
Epoch [11/50], Iter [75/209] Loss: 3.4157, average_loss: 3.5691
Epoch [11/50], Iter [80/209] Loss: 3.3971, average_loss: 3.5477
Epoch [11/50], Iter [85/209] Loss: 2.9404, average_loss: 3.5370
Epoch [11/50], Iter [90/209] Loss: 3.3924, average_loss: 3.5664
Epoch [11/50], Iter [95/209] Loss: 3.0866, average_loss: 3.5512
Epoch [11/50], Iter [100/209] Loss: 3.7641, average_loss: 3.5486
Epoch [11/50], Iter [105/209] Loss: 3.3680, average_loss: 3.5324
Epoch [11/50], Iter [110/209] Loss: 3.0760, average_loss: 3.5363
Epoch [11/50], Iter [115/209] Loss: 3.6305, average_loss: 3.5265
Epoch [11/50], Iter [120/209] Loss: 2.7418, average_loss: 3.5450
Epoch [11/50], Iter [125/209] Loss: 2.9856, average_loss: 3.5581
Epoch [11/50], Iter [130/209] Loss: 3.9402, average_loss: 3.5682
Epoch [11/50], Iter [135/209] Loss: 3.1571, average_loss: 3.5681
Epoch [11/50], Iter [140/209] Loss: 3.5825, average_loss: 3.5791
Epoch [11/50], Iter [145/209] Loss: 3.7013, average_loss: 3.5820
Epoch [11/50], Iter [150/209] Loss: 2.6925, average_loss: 3.5797
Epoch [11/50], Iter [155/209] Loss: 4.4350, average_loss: 3.5921
Epoch [11/50], Iter [160/209] Loss: 3.3538, average_loss: 3.5757
Epoch [11/50], Iter [165/209] Loss: 2.4402, average_loss: 3.5748
Epoch [11/50], Iter [170/209] Loss: 3.4291, average_loss: 3.5623
Epoch [11/50], Iter [175/209] Loss: 3.4791, average_loss: 3.5726
Epoch [11/50], Iter [180/209] Loss: 3.8178, average_loss: 3.5727
Epoch [11/50], Iter [185/209] Loss: 3.3872, average_loss: 3.5677
Epoch [11/50], Iter [190/209] Loss: 3.4350, average_loss: 3.5728
Epoch [11/50], Iter [195/209] Loss: 3.5755, average_loss: 3.5695
Epoch [11/50], Iter [200/209] Loss: 3.5576, average_loss: 3.5707
Epoch [11/50], Iter [205/209] Loss: 3.2623, average_loss: 3.5660
Updating best test loss: 3.90307

Starting epoch 12 / 50

Learning Rate for this epoch: 0.001

Epoch [12/50], Iter [5/209] Loss: 2.7476, average_loss: 3.9162
Epoch [12/50], Iter [10/209] Loss: 3.7883, average_loss: 3.5613
Epoch [12/50], Iter [15/209] Loss: 3.2421, average_loss: 3.5842
Epoch [12/50], Iter [20/209] Loss: 2.4553, average_loss: 3.4790
Epoch [12/50], Iter [25/209] Loss: 3.0966, average_loss: 3.4112
Epoch [12/50], Iter [30/209] Loss: 3.1337, average_loss: 3.4631
Epoch [12/50], Iter [35/209] Loss: 4.1297, average_loss: 3.5074
Epoch [12/50], Iter [40/209] Loss: 3.3088, average_loss: 3.5016
Epoch [12/50], Iter [45/209] Loss: 3.7152, average_loss: 3.4795
Epoch [12/50], Iter [50/209] Loss: 2.6294, average_loss: 3.4694
Epoch [12/50], Iter [55/209] Loss: 3.7205, average_loss: 3.4177
Epoch [12/50], Iter [60/209] Loss: 4.9971, average_loss: 3.4334

Epoch [12/50], Iter [65/209] Loss: 3.6641, average_loss: 3.4600
Epoch [12/50], Iter [70/209] Loss: 3.0617, average_loss: 3.4498
Epoch [12/50], Iter [75/209] Loss: 2.7310, average_loss: 3.4392
Epoch [12/50], Iter [80/209] Loss: 2.8883, average_loss: 3.4229
Epoch [12/50], Iter [85/209] Loss: 4.3957, average_loss: 3.4581
Epoch [12/50], Iter [90/209] Loss: 2.9027, average_loss: 3.4673
Epoch [12/50], Iter [95/209] Loss: 4.8750, average_loss: 3.5002
Epoch [12/50], Iter [100/209] Loss: 3.6009, average_loss: 3.5010
Epoch [12/50], Iter [105/209] Loss: 3.0816, average_loss: 3.4772
Epoch [12/50], Iter [110/209] Loss: 3.0980, average_loss: 3.4854
Epoch [12/50], Iter [115/209] Loss: 3.8180, average_loss: 3.4973
Epoch [12/50], Iter [120/209] Loss: 3.8585, average_loss: 3.5001
Epoch [12/50], Iter [125/209] Loss: 2.9116, average_loss: 3.4958
Epoch [12/50], Iter [130/209] Loss: 3.6816, average_loss: 3.5035
Epoch [12/50], Iter [135/209] Loss: 5.0295, average_loss: 3.5340
Epoch [12/50], Iter [140/209] Loss: 3.0031, average_loss: 3.5240
Epoch [12/50], Iter [145/209] Loss: 3.2508, average_loss: 3.5179
Epoch [12/50], Iter [150/209] Loss: 4.8591, average_loss: 3.5243
Epoch [12/50], Iter [155/209] Loss: 2.7659, average_loss: 3.5041
Epoch [12/50], Iter [160/209] Loss: 4.1768, average_loss: 3.4990
Epoch [12/50], Iter [165/209] Loss: 3.6593, average_loss: 3.5054
Epoch [12/50], Iter [170/209] Loss: 3.3492, average_loss: 3.4986
Epoch [12/50], Iter [175/209] Loss: 3.1201, average_loss: 3.4944
Epoch [12/50], Iter [180/209] Loss: 3.0787, average_loss: 3.4911
Epoch [12/50], Iter [185/209] Loss: 3.3116, average_loss: 3.4845
Epoch [12/50], Iter [190/209] Loss: 4.1822, average_loss: 3.4853
Epoch [12/50], Iter [195/209] Loss: 3.1201, average_loss: 3.4905
Epoch [12/50], Iter [200/209] Loss: 3.5692, average_loss: 3.4947
Epoch [12/50], Iter [205/209] Loss: 2.8779, average_loss: 3.4891
Updating best test loss: 3.86333

Starting epoch 13 / 50

Learning Rate for this epoch: 0.001

Epoch [13/50], Iter [5/209] Loss: 3.7035, average_loss: 3.7293
Epoch [13/50], Iter [10/209] Loss: 3.0410, average_loss: 3.4260
Epoch [13/50], Iter [15/209] Loss: 2.9967, average_loss: 3.3268
Epoch [13/50], Iter [20/209] Loss: 3.6253, average_loss: 3.3527
Epoch [13/50], Iter [25/209] Loss: 4.6075, average_loss: 3.4861
Epoch [13/50], Iter [30/209] Loss: 3.4919, average_loss: 3.5005
Epoch [13/50], Iter [35/209] Loss: 2.8278, average_loss: 3.5280
Epoch [13/50], Iter [40/209] Loss: 3.0736, average_loss: 3.5503
Epoch [13/50], Iter [45/209] Loss: 2.8762, average_loss: 3.4872
Epoch [13/50], Iter [50/209] Loss: 2.6258, average_loss: 3.4914
Epoch [13/50], Iter [55/209] Loss: 4.6945, average_loss: 3.5126
Epoch [13/50], Iter [60/209] Loss: 3.6330, average_loss: 3.4980
Epoch [13/50], Iter [65/209] Loss: 2.8453, average_loss: 3.4976
Epoch [13/50], Iter [70/209] Loss: 3.1060, average_loss: 3.4773
Epoch [13/50], Iter [75/209] Loss: 3.5681, average_loss: 3.4704
Epoch [13/50], Iter [80/209] Loss: 4.2389, average_loss: 3.4894
Epoch [13/50], Iter [85/209] Loss: 2.9520, average_loss: 3.4716
Epoch [13/50], Iter [90/209] Loss: 3.2699, average_loss: 3.4591
Epoch [13/50], Iter [95/209] Loss: 2.6039, average_loss: 3.4380
Epoch [13/50], Iter [100/209] Loss: 2.9042, average_loss: 3.4227

Epoch [13/50], Iter [105/209] Loss: 3.4137, average_loss: 3.4192
Epoch [13/50], Iter [110/209] Loss: 4.1537, average_loss: 3.4277
Epoch [13/50], Iter [115/209] Loss: 3.8717, average_loss: 3.4345
Epoch [13/50], Iter [120/209] Loss: 2.9245, average_loss: 3.4240
Epoch [13/50], Iter [125/209] Loss: 3.1535, average_loss: 3.4028
Epoch [13/50], Iter [130/209] Loss: 4.2672, average_loss: 3.4151
Epoch [13/50], Iter [135/209] Loss: 3.2593, average_loss: 3.4104
Epoch [13/50], Iter [140/209] Loss: 3.9874, average_loss: 3.4124
Epoch [13/50], Iter [145/209] Loss: 3.5895, average_loss: 3.4147
Epoch [13/50], Iter [150/209] Loss: 4.8736, average_loss: 3.4238
Epoch [13/50], Iter [155/209] Loss: 3.7540, average_loss: 3.4291
Epoch [13/50], Iter [160/209] Loss: 3.7208, average_loss: 3.4343
Epoch [13/50], Iter [165/209] Loss: 4.0138, average_loss: 3.4328
Epoch [13/50], Iter [170/209] Loss: 3.3909, average_loss: 3.4430
Epoch [13/50], Iter [175/209] Loss: 4.4587, average_loss: 3.4418
Epoch [13/50], Iter [180/209] Loss: 3.6511, average_loss: 3.4376
Epoch [13/50], Iter [185/209] Loss: 2.6435, average_loss: 3.4357
Epoch [13/50], Iter [190/209] Loss: 2.8223, average_loss: 3.4296
Epoch [13/50], Iter [195/209] Loss: 2.8534, average_loss: 3.4328
Epoch [13/50], Iter [200/209] Loss: 3.0950, average_loss: 3.4174
Epoch [13/50], Iter [205/209] Loss: 3.4992, average_loss: 3.4064
Updating best test loss: 3.81143

Starting epoch 14 / 50
Learning Rate for this epoch: 0.001
Epoch [14/50], Iter [5/209] Loss: 2.9489, average_loss: 3.7692
Epoch [14/50], Iter [10/209] Loss: 3.9660, average_loss: 3.6388
Epoch [14/50], Iter [15/209] Loss: 2.6419, average_loss: 3.5554
Epoch [14/50], Iter [20/209] Loss: 2.5318, average_loss: 3.4077
Epoch [14/50], Iter [25/209] Loss: 3.1138, average_loss: 3.4137
Epoch [14/50], Iter [30/209] Loss: 2.0919, average_loss: 3.3578
Epoch [14/50], Iter [35/209] Loss: 3.1872, average_loss: 3.2757
Epoch [14/50], Iter [40/209] Loss: 2.3753, average_loss: 3.2753
Epoch [14/50], Iter [45/209] Loss: 3.1486, average_loss: 3.3032
Epoch [14/50], Iter [50/209] Loss: 2.4232, average_loss: 3.3173
Epoch [14/50], Iter [55/209] Loss: 3.5147, average_loss: 3.3211
Epoch [14/50], Iter [60/209] Loss: 3.7626, average_loss: 3.3579
Epoch [14/50], Iter [65/209] Loss: 4.4982, average_loss: 3.3836
Epoch [14/50], Iter [70/209] Loss: 3.2370, average_loss: 3.3815
Epoch [14/50], Iter [75/209] Loss: 3.2394, average_loss: 3.3825
Epoch [14/50], Iter [80/209] Loss: 3.5697, average_loss: 3.3849
Epoch [14/50], Iter [85/209] Loss: 2.8198, average_loss: 3.3875
Epoch [14/50], Iter [90/209] Loss: 2.4997, average_loss: 3.3809
Epoch [14/50], Iter [95/209] Loss: 3.2167, average_loss: 3.3598
Epoch [14/50], Iter [100/209] Loss: 4.1814, average_loss: 3.3663
Epoch [14/50], Iter [105/209] Loss: 2.7944, average_loss: 3.3740
Epoch [14/50], Iter [110/209] Loss: 4.0718, average_loss: 3.3655
Epoch [14/50], Iter [115/209] Loss: 3.1708, average_loss: 3.3794
Epoch [14/50], Iter [120/209] Loss: 2.9260, average_loss: 3.3819
Epoch [14/50], Iter [125/209] Loss: 3.4771, average_loss: 3.3960
Epoch [14/50], Iter [130/209] Loss: 4.5515, average_loss: 3.3940
Epoch [14/50], Iter [135/209] Loss: 2.8476, average_loss: 3.3910
Epoch [14/50], Iter [140/209] Loss: 4.4818, average_loss: 3.3938

Epoch [14/50], Iter [145/209] Loss: 2.8190, average_loss: 3.3788
Epoch [14/50], Iter [150/209] Loss: 2.7036, average_loss: 3.3690
Epoch [14/50], Iter [155/209] Loss: 3.3385, average_loss: 3.3610
Epoch [14/50], Iter [160/209] Loss: 2.9505, average_loss: 3.3541
Epoch [14/50], Iter [165/209] Loss: 3.4344, average_loss: 3.3505
Epoch [14/50], Iter [170/209] Loss: 3.5608, average_loss: 3.3433
Epoch [14/50], Iter [175/209] Loss: 2.9389, average_loss: 3.3341
Epoch [14/50], Iter [180/209] Loss: 3.6780, average_loss: 3.3322
Epoch [14/50], Iter [185/209] Loss: 3.5558, average_loss: 3.3418
Epoch [14/50], Iter [190/209] Loss: 3.9562, average_loss: 3.3448
Epoch [14/50], Iter [195/209] Loss: 3.3808, average_loss: 3.3467
Epoch [14/50], Iter [200/209] Loss: 2.6286, average_loss: 3.3345
Epoch [14/50], Iter [205/209] Loss: 2.9963, average_loss: 3.3278
Updating best test loss: 3.76457

Starting epoch 15 / 50
Learning Rate for this epoch: 0.001
Epoch [15/50], Iter [5/209] Loss: 2.7861, average_loss: 3.0296
Epoch [15/50], Iter [10/209] Loss: 2.5658, average_loss: 3.1693
Epoch [15/50], Iter [15/209] Loss: 2.7234, average_loss: 2.9880
Epoch [15/50], Iter [20/209] Loss: 2.6575, average_loss: 3.0412
Epoch [15/50], Iter [25/209] Loss: 4.0064, average_loss: 3.1135
Epoch [15/50], Iter [30/209] Loss: 2.8821, average_loss: 3.1703
Epoch [15/50], Iter [35/209] Loss: 2.6604, average_loss: 3.1787
Epoch [15/50], Iter [40/209] Loss: 2.9103, average_loss: 3.1642
Epoch [15/50], Iter [45/209] Loss: 3.3692, average_loss: 3.1935
Epoch [15/50], Iter [50/209] Loss: 2.3932, average_loss: 3.1902
Epoch [15/50], Iter [55/209] Loss: 2.1544, average_loss: 3.1864
Epoch [15/50], Iter [60/209] Loss: 4.3109, average_loss: 3.2419
Epoch [15/50], Iter [65/209] Loss: 4.0135, average_loss: 3.2593
Epoch [15/50], Iter [70/209] Loss: 3.2451, average_loss: 3.2807
Epoch [15/50], Iter [75/209] Loss: 3.5149, average_loss: 3.2959
Epoch [15/50], Iter [80/209] Loss: 2.7868, average_loss: 3.3024
Epoch [15/50], Iter [85/209] Loss: 3.5007, average_loss: 3.3042
Epoch [15/50], Iter [90/209] Loss: 3.0045, average_loss: 3.2962
Epoch [15/50], Iter [95/209] Loss: 3.7893, average_loss: 3.3012
Epoch [15/50], Iter [100/209] Loss: 2.5514, average_loss: 3.2956
Epoch [15/50], Iter [105/209] Loss: 2.8015, average_loss: 3.2989
Epoch [15/50], Iter [110/209] Loss: 3.8053, average_loss: 3.2885
Epoch [15/50], Iter [115/209] Loss: 2.6558, average_loss: 3.2789
Epoch [15/50], Iter [120/209] Loss: 2.4269, average_loss: 3.2758
Epoch [15/50], Iter [125/209] Loss: 3.3769, average_loss: 3.2698
Epoch [15/50], Iter [130/209] Loss: 3.5008, average_loss: 3.2690
Epoch [15/50], Iter [135/209] Loss: 2.8945, average_loss: 3.2606
Epoch [15/50], Iter [140/209] Loss: 3.4821, average_loss: 3.2653
Epoch [15/50], Iter [145/209] Loss: 3.6068, average_loss: 3.2662
Epoch [15/50], Iter [150/209] Loss: 3.6194, average_loss: 3.2720
Epoch [15/50], Iter [155/209] Loss: 2.9632, average_loss: 3.2628
Epoch [15/50], Iter [160/209] Loss: 3.2763, average_loss: 3.2590
Epoch [15/50], Iter [165/209] Loss: 3.3380, average_loss: 3.2595
Epoch [15/50], Iter [170/209] Loss: 4.0991, average_loss: 3.2626
Epoch [15/50], Iter [175/209] Loss: 2.8499, average_loss: 3.2680
Epoch [15/50], Iter [180/209] Loss: 3.4018, average_loss: 3.2654

Epoch [15/50], Iter [185/209] Loss: 4.2692, average_loss: 3.2743
Epoch [15/50], Iter [190/209] Loss: 2.9735, average_loss: 3.2703
Epoch [15/50], Iter [195/209] Loss: 3.1688, average_loss: 3.2580
Epoch [15/50], Iter [200/209] Loss: 3.9401, average_loss: 3.2587
Epoch [15/50], Iter [205/209] Loss: 4.5213, average_loss: 3.2568

Starting epoch 16 / 50
Learning Rate for this epoch: 0.001
Epoch [16/50], Iter [5/209] Loss: 3.0518, average_loss: 2.8275
Epoch [16/50], Iter [10/209] Loss: 3.0988, average_loss: 2.9119
Epoch [16/50], Iter [15/209] Loss: 3.1688, average_loss: 3.1421
Epoch [16/50], Iter [20/209] Loss: 2.6291, average_loss: 3.1137
Epoch [16/50], Iter [25/209] Loss: 4.5581, average_loss: 3.1356
Epoch [16/50], Iter [30/209] Loss: 3.0203, average_loss: 3.0853
Epoch [16/50], Iter [35/209] Loss: 3.1896, average_loss: 3.0798
Epoch [16/50], Iter [40/209] Loss: 2.8795, average_loss: 3.1175
Epoch [16/50], Iter [45/209] Loss: 3.1198, average_loss: 3.1171
Epoch [16/50], Iter [50/209] Loss: 4.1450, average_loss: 3.1140
Epoch [16/50], Iter [55/209] Loss: 2.9287, average_loss: 3.1153
Epoch [16/50], Iter [60/209] Loss: 2.4672, average_loss: 3.0883
Epoch [16/50], Iter [65/209] Loss: 3.2691, average_loss: 3.0894
Epoch [16/50], Iter [70/209] Loss: 3.3348, average_loss: 3.1119
Epoch [16/50], Iter [75/209] Loss: 3.0813, average_loss: 3.1150
Epoch [16/50], Iter [80/209] Loss: 3.1751, average_loss: 3.0996
Epoch [16/50], Iter [85/209] Loss: 3.3912, average_loss: 3.1002
Epoch [16/50], Iter [90/209] Loss: 3.5234, average_loss: 3.0990
Epoch [16/50], Iter [95/209] Loss: 3.6175, average_loss: 3.1078
Epoch [16/50], Iter [100/209] Loss: 4.8973, average_loss: 3.1319
Epoch [16/50], Iter [105/209] Loss: 3.2985, average_loss: 3.1206
Epoch [16/50], Iter [110/209] Loss: 3.9237, average_loss: 3.1261
Epoch [16/50], Iter [115/209] Loss: 3.1254, average_loss: 3.1497
Epoch [16/50], Iter [120/209] Loss: 2.4786, average_loss: 3.1510
Epoch [16/50], Iter [125/209] Loss: 4.8473, average_loss: 3.1783
Epoch [16/50], Iter [130/209] Loss: 3.3123, average_loss: 3.1759
Epoch [16/50], Iter [135/209] Loss: 2.3921, average_loss: 3.1560
Epoch [16/50], Iter [140/209] Loss: 3.2173, average_loss: 3.1475
Epoch [16/50], Iter [145/209] Loss: 3.0933, average_loss: 3.1446
Epoch [16/50], Iter [150/209] Loss: 2.4164, average_loss: 3.1449
Epoch [16/50], Iter [155/209] Loss: 4.0560, average_loss: 3.1591
Epoch [16/50], Iter [160/209] Loss: 3.1013, average_loss: 3.1621
Epoch [16/50], Iter [165/209] Loss: 3.2216, average_loss: 3.1886
Epoch [16/50], Iter [170/209] Loss: 3.3369, average_loss: 3.1854
Epoch [16/50], Iter [175/209] Loss: 3.2268, average_loss: 3.1987
Epoch [16/50], Iter [180/209] Loss: 4.0471, average_loss: 3.2105
Epoch [16/50], Iter [185/209] Loss: 3.6052, average_loss: 3.2212
Epoch [16/50], Iter [190/209] Loss: 2.5382, average_loss: 3.2202
Epoch [16/50], Iter [195/209] Loss: 2.4993, average_loss: 3.2128
Epoch [16/50], Iter [200/209] Loss: 3.2028, average_loss: 3.2237
Epoch [16/50], Iter [205/209] Loss: 2.9013, average_loss: 3.2246

Starting epoch 17 / 50
Learning Rate for this epoch: 0.001

Epoch [17/50], Iter [5/209] Loss: 2.8667, average_loss: 3.3036
Epoch [17/50], Iter [10/209] Loss: 2.9947, average_loss: 3.4437
Epoch [17/50], Iter [15/209] Loss: 3.5275, average_loss: 3.2740
Epoch [17/50], Iter [20/209] Loss: 2.8790, average_loss: 3.1903
Epoch [17/50], Iter [25/209] Loss: 3.0598, average_loss: 3.1821
Epoch [17/50], Iter [30/209] Loss: 2.7526, average_loss: 3.1625
Epoch [17/50], Iter [35/209] Loss: 2.4742, average_loss: 3.1748
Epoch [17/50], Iter [40/209] Loss: 2.5901, average_loss: 3.1781
Epoch [17/50], Iter [45/209] Loss: 3.0578, average_loss: 3.1666
Epoch [17/50], Iter [50/209] Loss: 2.0744, average_loss: 3.1436
Epoch [17/50], Iter [55/209] Loss: 3.3373, average_loss: 3.1430
Epoch [17/50], Iter [60/209] Loss: 3.8421, average_loss: 3.1586
Epoch [17/50], Iter [65/209] Loss: 3.0780, average_loss: 3.1413
Epoch [17/50], Iter [70/209] Loss: 3.5394, average_loss: 3.1485
Epoch [17/50], Iter [75/209] Loss: 3.7852, average_loss: 3.1651
Epoch [17/50], Iter [80/209] Loss: 2.5038, average_loss: 3.1901
Epoch [17/50], Iter [85/209] Loss: 3.7815, average_loss: 3.1952
Epoch [17/50], Iter [90/209] Loss: 4.2849, average_loss: 3.1941
Epoch [17/50], Iter [95/209] Loss: 2.9972, average_loss: 3.1812
Epoch [17/50], Iter [100/209] Loss: 2.8973, average_loss: 3.1854
Epoch [17/50], Iter [105/209] Loss: 3.1079, average_loss: 3.2035
Epoch [17/50], Iter [110/209] Loss: 3.5218, average_loss: 3.1907
Epoch [17/50], Iter [115/209] Loss: 3.3135, average_loss: 3.1792
Epoch [17/50], Iter [120/209] Loss: 3.0352, average_loss: 3.1898
Epoch [17/50], Iter [125/209] Loss: 2.7914, average_loss: 3.1693
Epoch [17/50], Iter [130/209] Loss: 2.9890, average_loss: 3.1735
Epoch [17/50], Iter [135/209] Loss: 4.6839, average_loss: 3.1800
Epoch [17/50], Iter [140/209] Loss: 3.2970, average_loss: 3.1877
Epoch [17/50], Iter [145/209] Loss: 3.1724, average_loss: 3.1823
Epoch [17/50], Iter [150/209] Loss: 2.3438, average_loss: 3.1787
Epoch [17/50], Iter [155/209] Loss: 2.7584, average_loss: 3.1809
Epoch [17/50], Iter [160/209] Loss: 2.5008, average_loss: 3.1756
Epoch [17/50], Iter [165/209] Loss: 3.1869, average_loss: 3.1804
Epoch [17/50], Iter [170/209] Loss: 3.1114, average_loss: 3.1759
Epoch [17/50], Iter [175/209] Loss: 3.9608, average_loss: 3.1827
Epoch [17/50], Iter [180/209] Loss: 2.9594, average_loss: 3.1729
Epoch [17/50], Iter [185/209] Loss: 2.8216, average_loss: 3.1687
Epoch [17/50], Iter [190/209] Loss: 3.6033, average_loss: 3.1703
Epoch [17/50], Iter [195/209] Loss: 3.0596, average_loss: 3.1780
Epoch [17/50], Iter [200/209] Loss: 2.7515, average_loss: 3.1701
Epoch [17/50], Iter [205/209] Loss: 3.5640, average_loss: 3.1680
Updating best test loss: 3.70521

Starting epoch 18 / 50

Learning Rate for this epoch: 0.001

Epoch [18/50], Iter [5/209] Loss: 3.3391, average_loss: 3.2534
Epoch [18/50], Iter [10/209] Loss: 3.5439, average_loss: 3.2523
Epoch [18/50], Iter [15/209] Loss: 2.9893, average_loss: 3.1602
Epoch [18/50], Iter [20/209] Loss: 2.9572, average_loss: 3.0971
Epoch [18/50], Iter [25/209] Loss: 2.5886, average_loss: 3.0366
Epoch [18/50], Iter [30/209] Loss: 2.6999, average_loss: 2.9856
Epoch [18/50], Iter [35/209] Loss: 3.0694, average_loss: 2.9664
Epoch [18/50], Iter [40/209] Loss: 3.8138, average_loss: 2.9651

Epoch [18/50], Iter [45/209] Loss: 3.4153, average_loss: 2.9698
Epoch [18/50], Iter [50/209] Loss: 3.6640, average_loss: 3.0039
Epoch [18/50], Iter [55/209] Loss: 3.6808, average_loss: 2.9749
Epoch [18/50], Iter [60/209] Loss: 2.9715, average_loss: 3.0100
Epoch [18/50], Iter [65/209] Loss: 3.1590, average_loss: 3.0323
Epoch [18/50], Iter [70/209] Loss: 2.5307, average_loss: 3.0333
Epoch [18/50], Iter [75/209] Loss: 3.4607, average_loss: 3.0437
Epoch [18/50], Iter [80/209] Loss: 3.1091, average_loss: 3.0328
Epoch [18/50], Iter [85/209] Loss: 3.3649, average_loss: 3.0338
Epoch [18/50], Iter [90/209] Loss: 3.1174, average_loss: 3.0427
Epoch [18/50], Iter [95/209] Loss: 3.6981, average_loss: 3.0314
Epoch [18/50], Iter [100/209] Loss: 4.1086, average_loss: 3.0455
Epoch [18/50], Iter [105/209] Loss: 3.6001, average_loss: 3.0699
Epoch [18/50], Iter [110/209] Loss: 2.8686, average_loss: 3.0812
Epoch [18/50], Iter [115/209] Loss: 2.9444, average_loss: 3.1069
Epoch [18/50], Iter [120/209] Loss: 3.1175, average_loss: 3.0895
Epoch [18/50], Iter [125/209] Loss: 3.7663, average_loss: 3.0971
Epoch [18/50], Iter [130/209] Loss: 3.0837, average_loss: 3.0962
Epoch [18/50], Iter [135/209] Loss: 3.3243, average_loss: 3.0920
Epoch [18/50], Iter [140/209] Loss: 2.5673, average_loss: 3.0842
Epoch [18/50], Iter [145/209] Loss: 3.0850, average_loss: 3.0910
Epoch [18/50], Iter [150/209] Loss: 3.1433, average_loss: 3.0873
Epoch [18/50], Iter [155/209] Loss: 2.5027, average_loss: 3.0829
Epoch [18/50], Iter [160/209] Loss: 3.3194, average_loss: 3.0903
Epoch [18/50], Iter [165/209] Loss: 3.5230, average_loss: 3.0901
Epoch [18/50], Iter [170/209] Loss: 2.9605, average_loss: 3.0820
Epoch [18/50], Iter [175/209] Loss: 3.3217, average_loss: 3.0868
Epoch [18/50], Iter [180/209] Loss: 2.6744, average_loss: 3.0733
Epoch [18/50], Iter [185/209] Loss: 2.9238, average_loss: 3.0793
Epoch [18/50], Iter [190/209] Loss: 2.5057, average_loss: 3.0769
Epoch [18/50], Iter [195/209] Loss: 3.3756, average_loss: 3.0739
Epoch [18/50], Iter [200/209] Loss: 3.0208, average_loss: 3.0765
Epoch [18/50], Iter [205/209] Loss: 4.6578, average_loss: 3.0777
Updating best test loss: 3.66320

Starting epoch 19 / 50

Learning Rate for this epoch: 0.001

Epoch [19/50], Iter [5/209] Loss: 3.2155, average_loss: 3.1833
Epoch [19/50], Iter [10/209] Loss: 2.5580, average_loss: 2.9900
Epoch [19/50], Iter [15/209] Loss: 3.4150, average_loss: 2.9990
Epoch [19/50], Iter [20/209] Loss: 2.4326, average_loss: 3.0408
Epoch [19/50], Iter [25/209] Loss: 3.2819, average_loss: 3.0009
Epoch [19/50], Iter [30/209] Loss: 3.3568, average_loss: 3.0473
Epoch [19/50], Iter [35/209] Loss: 3.3728, average_loss: 3.0741
Epoch [19/50], Iter [40/209] Loss: 2.6235, average_loss: 3.0248
Epoch [19/50], Iter [45/209] Loss: 2.9505, average_loss: 3.0390
Epoch [19/50], Iter [50/209] Loss: 2.8072, average_loss: 3.0512
Epoch [19/50], Iter [55/209] Loss: 2.8816, average_loss: 3.0630
Epoch [19/50], Iter [60/209] Loss: 3.6177, average_loss: 3.0740
Epoch [19/50], Iter [65/209] Loss: 3.4869, average_loss: 3.0870
Epoch [19/50], Iter [70/209] Loss: 2.5934, average_loss: 3.0591
Epoch [19/50], Iter [75/209] Loss: 2.9460, average_loss: 3.0317
Epoch [19/50], Iter [80/209] Loss: 2.8291, average_loss: 3.0208

Epoch [19/50], Iter [85/209] Loss: 2.6921, average_loss: 3.0035
Epoch [19/50], Iter [90/209] Loss: 3.5252, average_loss: 2.9919
Epoch [19/50], Iter [95/209] Loss: 3.3949, average_loss: 3.0125
Epoch [19/50], Iter [100/209] Loss: 3.3464, average_loss: 3.0197
Epoch [19/50], Iter [105/209] Loss: 3.5187, average_loss: 3.0310
Epoch [19/50], Iter [110/209] Loss: 2.4517, average_loss: 3.0174
Epoch [19/50], Iter [115/209] Loss: 2.0572, average_loss: 2.9999
Epoch [19/50], Iter [120/209] Loss: 3.0738, average_loss: 2.9967
Epoch [19/50], Iter [125/209] Loss: 3.9892, average_loss: 3.0028
Epoch [19/50], Iter [130/209] Loss: 4.0748, average_loss: 3.0074
Epoch [19/50], Iter [135/209] Loss: 3.1290, average_loss: 3.0101
Epoch [19/50], Iter [140/209] Loss: 2.2526, average_loss: 3.0042
Epoch [19/50], Iter [145/209] Loss: 3.9142, average_loss: 3.0088
Epoch [19/50], Iter [150/209] Loss: 3.1062, average_loss: 3.0085
Epoch [19/50], Iter [155/209] Loss: 3.9056, average_loss: 3.0148
Epoch [19/50], Iter [160/209] Loss: 2.9296, average_loss: 3.0083
Epoch [19/50], Iter [165/209] Loss: 3.1097, average_loss: 3.0192
Epoch [19/50], Iter [170/209] Loss: 2.4274, average_loss: 3.0135
Epoch [19/50], Iter [175/209] Loss: 2.4664, average_loss: 3.0232
Epoch [19/50], Iter [180/209] Loss: 3.4339, average_loss: 3.0226
Epoch [19/50], Iter [185/209] Loss: 3.8252, average_loss: 3.0364
Epoch [19/50], Iter [190/209] Loss: 2.2646, average_loss: 3.0452
Epoch [19/50], Iter [195/209] Loss: 3.0388, average_loss: 3.0393
Epoch [19/50], Iter [200/209] Loss: 2.9863, average_loss: 3.0317
Epoch [19/50], Iter [205/209] Loss: 3.1545, average_loss: 3.0424

Starting epoch 20 / 50

Learning Rate for this epoch: 0.001

Epoch [20/50], Iter [5/209] Loss: 4.3322, average_loss: 3.2064
Epoch [20/50], Iter [10/209] Loss: 2.4343, average_loss: 2.8791
Epoch [20/50], Iter [15/209] Loss: 2.7812, average_loss: 2.9604
Epoch [20/50], Iter [20/209] Loss: 3.1100, average_loss: 2.9193
Epoch [20/50], Iter [25/209] Loss: 2.5254, average_loss: 2.9266
Epoch [20/50], Iter [30/209] Loss: 3.1105, average_loss: 2.9511
Epoch [20/50], Iter [35/209] Loss: 2.4127, average_loss: 2.9020
Epoch [20/50], Iter [40/209] Loss: 2.4424, average_loss: 2.9383
Epoch [20/50], Iter [45/209] Loss: 2.5459, average_loss: 2.9757
Epoch [20/50], Iter [50/209] Loss: 3.9410, average_loss: 2.9913
Epoch [20/50], Iter [55/209] Loss: 3.3595, average_loss: 3.0093
Epoch [20/50], Iter [60/209] Loss: 2.4245, average_loss: 2.9825
Epoch [20/50], Iter [65/209] Loss: 3.3852, average_loss: 2.9988
Epoch [20/50], Iter [70/209] Loss: 2.9607, average_loss: 3.0228
Epoch [20/50], Iter [75/209] Loss: 2.6687, average_loss: 3.0121
Epoch [20/50], Iter [80/209] Loss: 1.8600, average_loss: 2.9850
Epoch [20/50], Iter [85/209] Loss: 2.7943, average_loss: 2.9895
Epoch [20/50], Iter [90/209] Loss: 3.2871, average_loss: 2.9877
Epoch [20/50], Iter [95/209] Loss: 2.5315, average_loss: 2.9721
Epoch [20/50], Iter [100/209] Loss: 3.8470, average_loss: 2.9686
Epoch [20/50], Iter [105/209] Loss: 3.4211, average_loss: 2.9685
Epoch [20/50], Iter [110/209] Loss: 3.8094, average_loss: 2.9971
Epoch [20/50], Iter [115/209] Loss: 2.7515, average_loss: 2.9837
Epoch [20/50], Iter [120/209] Loss: 3.0706, average_loss: 2.9835
Epoch [20/50], Iter [125/209] Loss: 4.0847, average_loss: 2.9852

Epoch [20/50], Iter [130/209] Loss: 2.1717, average_loss: 2.9733
Epoch [20/50], Iter [135/209] Loss: 2.8637, average_loss: 2.9776
Epoch [20/50], Iter [140/209] Loss: 2.4212, average_loss: 2.9883
Epoch [20/50], Iter [145/209] Loss: 3.7586, average_loss: 3.0005
Epoch [20/50], Iter [150/209] Loss: 2.6609, average_loss: 2.9915
Epoch [20/50], Iter [155/209] Loss: 2.3517, average_loss: 3.0031
Epoch [20/50], Iter [160/209] Loss: 2.5879, average_loss: 3.0000
Epoch [20/50], Iter [165/209] Loss: 2.4645, average_loss: 3.0060
Epoch [20/50], Iter [170/209] Loss: 3.2548, average_loss: 3.0128
Epoch [20/50], Iter [175/209] Loss: 3.6683, average_loss: 3.0217
Epoch [20/50], Iter [180/209] Loss: 2.7696, average_loss: 3.0129
Epoch [20/50], Iter [185/209] Loss: 2.5239, average_loss: 3.0138
Epoch [20/50], Iter [190/209] Loss: 2.3735, average_loss: 3.0067
Epoch [20/50], Iter [195/209] Loss: 3.0567, average_loss: 2.9984
Epoch [20/50], Iter [200/209] Loss: 2.7298, average_loss: 2.9993
Epoch [20/50], Iter [205/209] Loss: 4.0106, average_loss: 3.0076

Starting epoch 21 / 50

Learning Rate for this epoch: 0.001

Epoch [21/50], Iter [5/209] Loss: 3.8573, average_loss: 3.0815
Epoch [21/50], Iter [10/209] Loss: 2.3201, average_loss: 2.8012
Epoch [21/50], Iter [15/209] Loss: 3.2308, average_loss: 2.8118
Epoch [21/50], Iter [20/209] Loss: 2.5980, average_loss: 2.8229
Epoch [21/50], Iter [25/209] Loss: 3.3424, average_loss: 2.8597
Epoch [21/50], Iter [30/209] Loss: 4.5426, average_loss: 2.9560
Epoch [21/50], Iter [35/209] Loss: 2.2445, average_loss: 2.9478
Epoch [21/50], Iter [40/209] Loss: 2.2426, average_loss: 2.8805
Epoch [21/50], Iter [45/209] Loss: 3.0474, average_loss: 2.8553
Epoch [21/50], Iter [50/209] Loss: 2.6773, average_loss: 2.8441
Epoch [21/50], Iter [55/209] Loss: 2.6610, average_loss: 2.8285
Epoch [21/50], Iter [60/209] Loss: 2.2857, average_loss: 2.7927
Epoch [21/50], Iter [65/209] Loss: 3.4319, average_loss: 2.8060
Epoch [21/50], Iter [70/209] Loss: 3.2128, average_loss: 2.8313
Epoch [21/50], Iter [75/209] Loss: 2.1042, average_loss: 2.7947
Epoch [21/50], Iter [80/209] Loss: 2.8788, average_loss: 2.8156
Epoch [21/50], Iter [85/209] Loss: 4.3289, average_loss: 2.8457
Epoch [21/50], Iter [90/209] Loss: 2.7303, average_loss: 2.8558
Epoch [21/50], Iter [95/209] Loss: 2.7663, average_loss: 2.8541
Epoch [21/50], Iter [100/209] Loss: 3.3608, average_loss: 2.8537
Epoch [21/50], Iter [105/209] Loss: 3.0048, average_loss: 2.8658
Epoch [21/50], Iter [110/209] Loss: 3.6661, average_loss: 2.8801
Epoch [21/50], Iter [115/209] Loss: 2.8910, average_loss: 2.9055
Epoch [21/50], Iter [120/209] Loss: 3.3273, average_loss: 2.9071
Epoch [21/50], Iter [125/209] Loss: 2.6903, average_loss: 2.9062
Epoch [21/50], Iter [130/209] Loss: 3.5680, average_loss: 2.9115
Epoch [21/50], Iter [135/209] Loss: 2.9637, average_loss: 2.9470
Epoch [21/50], Iter [140/209] Loss: 2.7348, average_loss: 2.9455
Epoch [21/50], Iter [145/209] Loss: 2.9940, average_loss: 2.9382
Epoch [21/50], Iter [150/209] Loss: 3.0983, average_loss: 2.9362
Epoch [21/50], Iter [155/209] Loss: 4.1174, average_loss: 2.9423
Epoch [21/50], Iter [160/209] Loss: 2.8081, average_loss: 2.9464
Epoch [21/50], Iter [165/209] Loss: 2.3117, average_loss: 2.9413
Epoch [21/50], Iter [170/209] Loss: 2.6753, average_loss: 2.9445

Epoch [21/50], Iter [175/209] Loss: 3.3648, average_loss: 2.9465
Epoch [21/50], Iter [180/209] Loss: 3.1534, average_loss: 2.9469
Epoch [21/50], Iter [185/209] Loss: 3.1582, average_loss: 2.9609
Epoch [21/50], Iter [190/209] Loss: 4.2259, average_loss: 2.9580
Epoch [21/50], Iter [195/209] Loss: 3.6049, average_loss: 2.9609
Epoch [21/50], Iter [200/209] Loss: 3.3542, average_loss: 2.9656
Epoch [21/50], Iter [205/209] Loss: 3.1990, average_loss: 2.9644

Starting epoch 22 / 50

Learning Rate for this epoch: 0.001

Epoch [22/50], Iter [5/209] Loss: 2.8358, average_loss: 2.9455
Epoch [22/50], Iter [10/209] Loss: 3.0969, average_loss: 3.0230
Epoch [22/50], Iter [15/209] Loss: 3.2819, average_loss: 2.9776
Epoch [22/50], Iter [20/209] Loss: 2.4778, average_loss: 2.9268
Epoch [22/50], Iter [25/209] Loss: 2.5757, average_loss: 2.9524
Epoch [22/50], Iter [30/209] Loss: 2.5873, average_loss: 2.9987
Epoch [22/50], Iter [35/209] Loss: 2.5354, average_loss: 2.9831
Epoch [22/50], Iter [40/209] Loss: 2.4463, average_loss: 2.9603
Epoch [22/50], Iter [45/209] Loss: 3.7602, average_loss: 2.9552
Epoch [22/50], Iter [50/209] Loss: 3.1098, average_loss: 2.9691
Epoch [22/50], Iter [55/209] Loss: 3.0503, average_loss: 2.9613
Epoch [22/50], Iter [60/209] Loss: 2.3939, average_loss: 2.9502
Epoch [22/50], Iter [65/209] Loss: 3.2811, average_loss: 2.9783
Epoch [22/50], Iter [70/209] Loss: 3.4456, average_loss: 2.9612
Epoch [22/50], Iter [75/209] Loss: 2.6317, average_loss: 2.9520
Epoch [22/50], Iter [80/209] Loss: 3.7523, average_loss: 2.9641
Epoch [22/50], Iter [85/209] Loss: 2.3982, average_loss: 2.9389
Epoch [22/50], Iter [90/209] Loss: 2.8862, average_loss: 2.9364
Epoch [22/50], Iter [95/209] Loss: 2.9080, average_loss: 2.9175
Epoch [22/50], Iter [100/209] Loss: 2.0295, average_loss: 2.9060
Epoch [22/50], Iter [105/209] Loss: 3.0581, average_loss: 2.9220
Epoch [22/50], Iter [110/209] Loss: 2.8037, average_loss: 2.9199
Epoch [22/50], Iter [115/209] Loss: 3.3845, average_loss: 2.9302
Epoch [22/50], Iter [120/209] Loss: 2.7491, average_loss: 2.9367
Epoch [22/50], Iter [125/209] Loss: 3.3288, average_loss: 2.9317
Epoch [22/50], Iter [130/209] Loss: 2.2855, average_loss: 2.9203
Epoch [22/50], Iter [135/209] Loss: 2.6836, average_loss: 2.9168
Epoch [22/50], Iter [140/209] Loss: 3.1940, average_loss: 2.9246
Epoch [22/50], Iter [145/209] Loss: 2.2341, average_loss: 2.9308
Epoch [22/50], Iter [150/209] Loss: 4.1178, average_loss: 2.9466
Epoch [22/50], Iter [155/209] Loss: 2.9208, average_loss: 2.9353
Epoch [22/50], Iter [160/209] Loss: 2.4559, average_loss: 2.9430
Epoch [22/50], Iter [165/209] Loss: 2.8581, average_loss: 2.9425
Epoch [22/50], Iter [170/209] Loss: 3.2482, average_loss: 2.9481
Epoch [22/50], Iter [175/209] Loss: 3.1070, average_loss: 2.9507
Epoch [22/50], Iter [180/209] Loss: 3.0107, average_loss: 2.9530
Epoch [22/50], Iter [185/209] Loss: 2.2357, average_loss: 2.9526
Epoch [22/50], Iter [190/209] Loss: 2.8283, average_loss: 2.9539
Epoch [22/50], Iter [195/209] Loss: 2.7239, average_loss: 2.9538
Epoch [22/50], Iter [200/209] Loss: 3.4683, average_loss: 2.9479
Epoch [22/50], Iter [205/209] Loss: 2.8316, average_loss: 2.9444

Updating best test loss: 3.62370

Starting epoch 23 / 50

Learning Rate for this epoch: 0.001

Epoch [23/50], Iter [5/209] Loss: 3.2607, average_loss: 2.9403
Epoch [23/50], Iter [10/209] Loss: 2.4774, average_loss: 2.9869
Epoch [23/50], Iter [15/209] Loss: 2.8252, average_loss: 2.8302
Epoch [23/50], Iter [20/209] Loss: 2.7054, average_loss: 2.8265
Epoch [23/50], Iter [25/209] Loss: 3.3638, average_loss: 2.8652
Epoch [23/50], Iter [30/209] Loss: 3.4159, average_loss: 2.8793
Epoch [23/50], Iter [35/209] Loss: 3.0288, average_loss: 2.8643
Epoch [23/50], Iter [40/209] Loss: 2.6319, average_loss: 2.8546
Epoch [23/50], Iter [45/209] Loss: 2.3714, average_loss: 2.8665
Epoch [23/50], Iter [50/209] Loss: 2.9490, average_loss: 2.8652
Epoch [23/50], Iter [55/209] Loss: 1.8037, average_loss: 2.8562
Epoch [23/50], Iter [60/209] Loss: 3.5437, average_loss: 2.8706
Epoch [23/50], Iter [65/209] Loss: 2.7764, average_loss: 2.8755
Epoch [23/50], Iter [70/209] Loss: 2.5289, average_loss: 2.9065
Epoch [23/50], Iter [75/209] Loss: 2.6550, average_loss: 2.9183
Epoch [23/50], Iter [80/209] Loss: 3.2554, average_loss: 2.9111
Epoch [23/50], Iter [85/209] Loss: 3.0175, average_loss: 2.8883
Epoch [23/50], Iter [90/209] Loss: 2.2310, average_loss: 2.8885
Epoch [23/50], Iter [95/209] Loss: 2.4679, average_loss: 2.8749
Epoch [23/50], Iter [100/209] Loss: 2.3954, average_loss: 2.8727
Epoch [23/50], Iter [105/209] Loss: 2.5558, average_loss: 2.8799
Epoch [23/50], Iter [110/209] Loss: 2.7029, average_loss: 2.8661
Epoch [23/50], Iter [115/209] Loss: 2.5929, average_loss: 2.8672
Epoch [23/50], Iter [120/209] Loss: 2.7315, average_loss: 2.8757
Epoch [23/50], Iter [125/209] Loss: 3.1959, average_loss: 2.8812
Epoch [23/50], Iter [130/209] Loss: 3.9205, average_loss: 2.8852
Epoch [23/50], Iter [135/209] Loss: 2.5180, average_loss: 2.8877
Epoch [23/50], Iter [140/209] Loss: 3.0741, average_loss: 2.8894
Epoch [23/50], Iter [145/209] Loss: 2.6539, average_loss: 2.8826
Epoch [23/50], Iter [150/209] Loss: 3.5167, average_loss: 2.8936
Epoch [23/50], Iter [155/209] Loss: 2.7632, average_loss: 2.9064
Epoch [23/50], Iter [160/209] Loss: 3.5573, average_loss: 2.9156
Epoch [23/50], Iter [165/209] Loss: 3.0014, average_loss: 2.9139
Epoch [23/50], Iter [170/209] Loss: 2.1721, average_loss: 2.9127
Epoch [23/50], Iter [175/209] Loss: 2.2535, average_loss: 2.9094
Epoch [23/50], Iter [180/209] Loss: 2.3270, average_loss: 2.9034
Epoch [23/50], Iter [185/209] Loss: 2.8002, average_loss: 2.9044
Epoch [23/50], Iter [190/209] Loss: 2.3081, average_loss: 2.9058
Epoch [23/50], Iter [195/209] Loss: 2.2672, average_loss: 2.9110
Epoch [23/50], Iter [200/209] Loss: 2.8042, average_loss: 2.9126
Epoch [23/50], Iter [205/209] Loss: 3.8059, average_loss: 2.9107

Starting epoch 24 / 50

Learning Rate for this epoch: 0.001

Epoch [24/50], Iter [5/209] Loss: 3.3936, average_loss: 2.9152
Epoch [24/50], Iter [10/209] Loss: 2.2254, average_loss: 2.9003
Epoch [24/50], Iter [15/209] Loss: 2.5361, average_loss: 2.9522
Epoch [24/50], Iter [20/209] Loss: 2.6289, average_loss: 2.9916
Epoch [24/50], Iter [25/209] Loss: 2.7408, average_loss: 2.9732
Epoch [24/50], Iter [30/209] Loss: 3.5505, average_loss: 3.0644

Epoch [24/50], Iter [35/209] Loss: 2.7587, average_loss: 3.0309
Epoch [24/50], Iter [40/209] Loss: 3.0970, average_loss: 3.0045
Epoch [24/50], Iter [45/209] Loss: 2.0428, average_loss: 3.0005
Epoch [24/50], Iter [50/209] Loss: 2.3447, average_loss: 2.9576
Epoch [24/50], Iter [55/209] Loss: 2.7289, average_loss: 2.9300
Epoch [24/50], Iter [60/209] Loss: 2.8664, average_loss: 2.9216
Epoch [24/50], Iter [65/209] Loss: 2.5942, average_loss: 2.9009
Epoch [24/50], Iter [70/209] Loss: 2.2948, average_loss: 2.8801
Epoch [24/50], Iter [75/209] Loss: 2.9896, average_loss: 2.8835
Epoch [24/50], Iter [80/209] Loss: 2.4578, average_loss: 2.8768
Epoch [24/50], Iter [85/209] Loss: 2.9068, average_loss: 2.8585
Epoch [24/50], Iter [90/209] Loss: 2.1575, average_loss: 2.8380
Epoch [24/50], Iter [95/209] Loss: 3.3913, average_loss: 2.8395
Epoch [24/50], Iter [100/209] Loss: 2.7991, average_loss: 2.8438
Epoch [24/50], Iter [105/209] Loss: 2.7802, average_loss: 2.8424
Epoch [24/50], Iter [110/209] Loss: 1.8493, average_loss: 2.8271
Epoch [24/50], Iter [115/209] Loss: 2.4952, average_loss: 2.8371
Epoch [24/50], Iter [120/209] Loss: 2.5958, average_loss: 2.8492
Epoch [24/50], Iter [125/209] Loss: 2.6842, average_loss: 2.8714
Epoch [24/50], Iter [130/209] Loss: 2.4896, average_loss: 2.8604
Epoch [24/50], Iter [135/209] Loss: 3.2080, average_loss: 2.8641
Epoch [24/50], Iter [140/209] Loss: 3.0548, average_loss: 2.8718
Epoch [24/50], Iter [145/209] Loss: 3.8952, average_loss: 2.8820
Epoch [24/50], Iter [150/209] Loss: 2.8482, average_loss: 2.8819
Epoch [24/50], Iter [155/209] Loss: 2.8347, average_loss: 2.8755
Epoch [24/50], Iter [160/209] Loss: 3.6443, average_loss: 2.8866
Epoch [24/50], Iter [165/209] Loss: 2.4204, average_loss: 2.8821
Epoch [24/50], Iter [170/209] Loss: 2.6100, average_loss: 2.8712
Epoch [24/50], Iter [175/209] Loss: 2.5662, average_loss: 2.8617
Epoch [24/50], Iter [180/209] Loss: 2.7568, average_loss: 2.8666
Epoch [24/50], Iter [185/209] Loss: 3.1096, average_loss: 2.8720
Epoch [24/50], Iter [190/209] Loss: 4.1357, average_loss: 2.8753
Epoch [24/50], Iter [195/209] Loss: 3.4110, average_loss: 2.8839
Epoch [24/50], Iter [200/209] Loss: 3.7553, average_loss: 2.8879
Epoch [24/50], Iter [205/209] Loss: 3.0381, average_loss: 2.8922

Starting epoch 25 / 50

Learning Rate for this epoch: 0.001

Epoch [25/50], Iter [5/209] Loss: 2.3990, average_loss: 2.5199
Epoch [25/50], Iter [10/209] Loss: 2.1090, average_loss: 2.6084
Epoch [25/50], Iter [15/209] Loss: 3.9536, average_loss: 2.6842
Epoch [25/50], Iter [20/209] Loss: 2.3350, average_loss: 2.7274
Epoch [25/50], Iter [25/209] Loss: 2.7796, average_loss: 2.7755
Epoch [25/50], Iter [30/209] Loss: 2.5058, average_loss: 2.8392
Epoch [25/50], Iter [35/209] Loss: 3.2889, average_loss: 2.9074
Epoch [25/50], Iter [40/209] Loss: 2.9442, average_loss: 2.9306
Epoch [25/50], Iter [45/209] Loss: 2.3454, average_loss: 2.8933
Epoch [25/50], Iter [50/209] Loss: 2.9507, average_loss: 2.8712
Epoch [25/50], Iter [55/209] Loss: 2.6531, average_loss: 2.9029
Epoch [25/50], Iter [60/209] Loss: 2.4822, average_loss: 2.8859
Epoch [25/50], Iter [65/209] Loss: 3.4330, average_loss: 2.8944
Epoch [25/50], Iter [70/209] Loss: 3.4716, average_loss: 2.9103
Epoch [25/50], Iter [75/209] Loss: 2.0101, average_loss: 2.8694

Epoch [25/50], Iter [80/209] Loss: 2.3889, average_loss: 2.8685
Epoch [25/50], Iter [85/209] Loss: 1.7299, average_loss: 2.8383
Epoch [25/50], Iter [90/209] Loss: 1.8780, average_loss: 2.8178
Epoch [25/50], Iter [95/209] Loss: 2.5389, average_loss: 2.8226
Epoch [25/50], Iter [100/209] Loss: 2.7794, average_loss: 2.8234
Epoch [25/50], Iter [105/209] Loss: 2.6125, average_loss: 2.8210
Epoch [25/50], Iter [110/209] Loss: 3.7547, average_loss: 2.8301
Epoch [25/50], Iter [115/209] Loss: 3.1881, average_loss: 2.8453
Epoch [25/50], Iter [120/209] Loss: 3.0767, average_loss: 2.8426
Epoch [25/50], Iter [125/209] Loss: 2.5385, average_loss: 2.8390
Epoch [25/50], Iter [130/209] Loss: 4.1524, average_loss: 2.8459
Epoch [25/50], Iter [135/209] Loss: 2.3731, average_loss: 2.8413
Epoch [25/50], Iter [140/209] Loss: 2.5383, average_loss: 2.8291
Epoch [25/50], Iter [145/209] Loss: 3.1890, average_loss: 2.8304
Epoch [25/50], Iter [150/209] Loss: 3.4367, average_loss: 2.8293
Epoch [25/50], Iter [155/209] Loss: 3.3144, average_loss: 2.8427
Epoch [25/50], Iter [160/209] Loss: 2.1872, average_loss: 2.8341
Epoch [25/50], Iter [165/209] Loss: 2.1064, average_loss: 2.8215
Epoch [25/50], Iter [170/209] Loss: 2.6451, average_loss: 2.8177
Epoch [25/50], Iter [175/209] Loss: 2.8802, average_loss: 2.8110
Epoch [25/50], Iter [180/209] Loss: 2.9505, average_loss: 2.8099
Epoch [25/50], Iter [185/209] Loss: 2.2587, average_loss: 2.8122
Epoch [25/50], Iter [190/209] Loss: 2.1616, average_loss: 2.7998
Epoch [25/50], Iter [195/209] Loss: 2.1820, average_loss: 2.8069
Epoch [25/50], Iter [200/209] Loss: 2.8548, average_loss: 2.8147
Epoch [25/50], Iter [205/209] Loss: 2.2633, average_loss: 2.8061
Updating best test loss: 3.59376

Starting epoch 26 / 50

Learning Rate for this epoch: 0.001

Epoch [26/50], Iter [5/209] Loss: 3.2294, average_loss: 2.9251
Epoch [26/50], Iter [10/209] Loss: 2.3812, average_loss: 2.7582
Epoch [26/50], Iter [15/209] Loss: 2.5405, average_loss: 2.7996
Epoch [26/50], Iter [20/209] Loss: 2.2859, average_loss: 2.8158
Epoch [26/50], Iter [25/209] Loss: 2.9789, average_loss: 2.8232
Epoch [26/50], Iter [30/209] Loss: 2.4489, average_loss: 2.8351
Epoch [26/50], Iter [35/209] Loss: 2.3343, average_loss: 2.8197
Epoch [26/50], Iter [40/209] Loss: 3.0699, average_loss: 2.8238
Epoch [26/50], Iter [45/209] Loss: 1.9060, average_loss: 2.7953
Epoch [26/50], Iter [50/209] Loss: 2.7531, average_loss: 2.7713
Epoch [26/50], Iter [55/209] Loss: 3.6594, average_loss: 2.7894
Epoch [26/50], Iter [60/209] Loss: 3.6302, average_loss: 2.7857
Epoch [26/50], Iter [65/209] Loss: 2.3453, average_loss: 2.7690
Epoch [26/50], Iter [70/209] Loss: 2.9967, average_loss: 2.7891
Epoch [26/50], Iter [75/209] Loss: 3.5027, average_loss: 2.7958
Epoch [26/50], Iter [80/209] Loss: 2.9489, average_loss: 2.7957
Epoch [26/50], Iter [85/209] Loss: 2.7206, average_loss: 2.7931
Epoch [26/50], Iter [90/209] Loss: 2.1387, average_loss: 2.8022
Epoch [26/50], Iter [95/209] Loss: 3.6691, average_loss: 2.8162
Epoch [26/50], Iter [100/209] Loss: 3.5049, average_loss: 2.8216
Epoch [26/50], Iter [105/209] Loss: 2.7683, average_loss: 2.8006
Epoch [26/50], Iter [110/209] Loss: 2.4163, average_loss: 2.7996
Epoch [26/50], Iter [115/209] Loss: 3.6810, average_loss: 2.8224

MP2_P2-3

March 30, 2020

```
[0]: 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
```

```
[0]: # from google.colab import drive
# drive.mount('/content/drive')
```

0.1 Initialization

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

```
[9]: print(device)
```

cuda:0

```
[0]: # 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.

```
[11]: load_network_path = "best_detector.pth"  
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)
```

Loading saved network from best_detector.pth

```
[0]: 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
```

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

0.2 Reading Pascal Data

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

The train dataset loader also using a variety of data augmentation techniques including random shift, scaling, crop, and flips. Data augmentation is slightly more complicated for detection dataset

since the bounding box annotations must be kept consistent through the transformations.

Since the output of the detector network we train is an $S \times S(B^*5+C)$, 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.

```
[14]: 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

```
[15]: 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

0.3 Train detector

```
[0]: # 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.
#     print("i")

#     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')

```

1 View example predictions

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

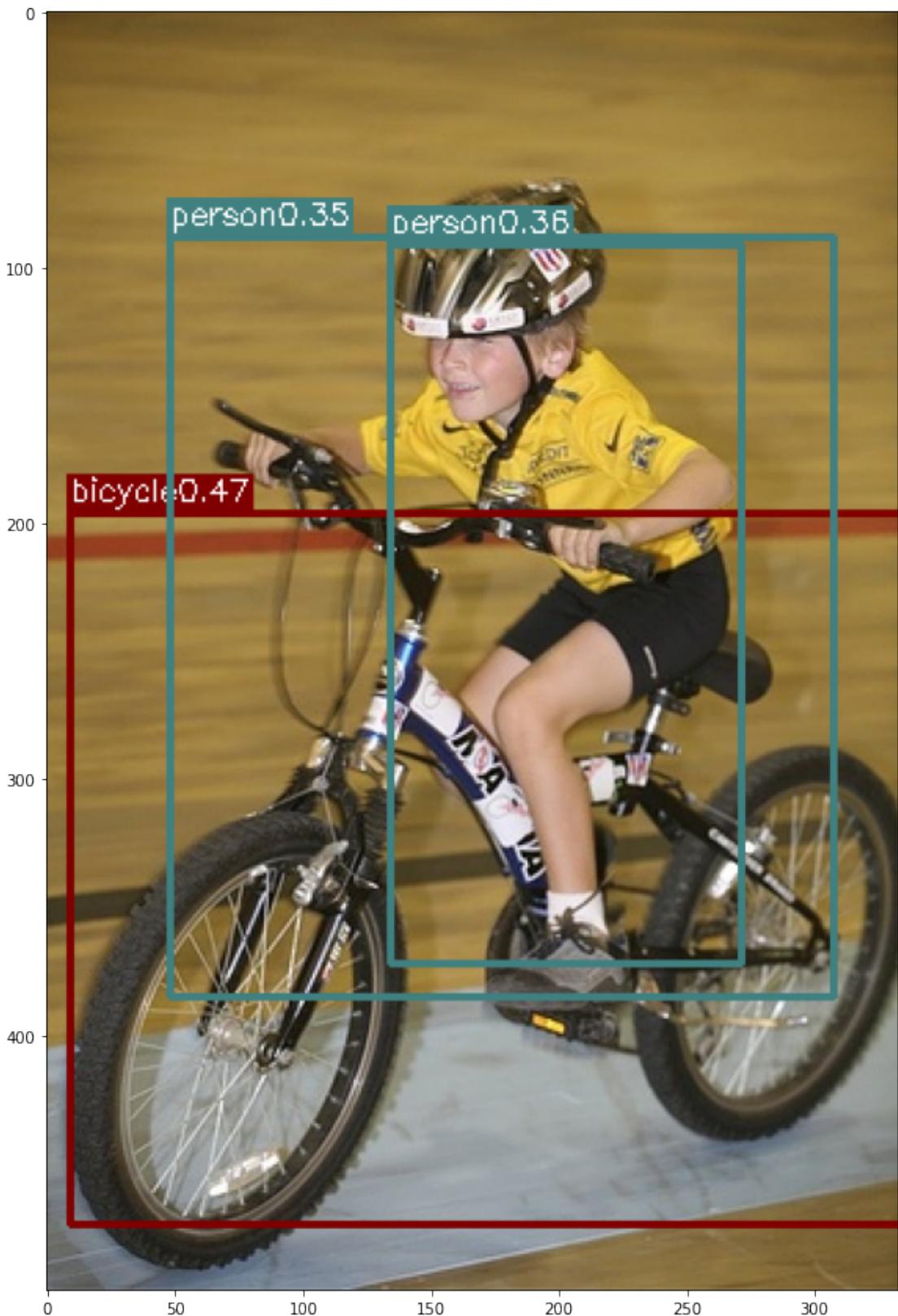
# select random image from test set
image_name = random.choice(test_dataset.fnames)
image = cv2.imread(os.path.join(file_root_test, image_name))
image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)

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...

```
[16]: <matplotlib.image.AxesImage at 0x7fa681f23eb8>
```



```
[17]: # !rmdir VOCdevkit  
# !rm VOCtest_06-Nov-2007.tar
```

```
rmdir: failed to remove 'VOCdevkit': No such file or directory  
rm: cannot remove 'VOCtest_06-Nov-2007.tar': No such file or directory
```

1.1 Evaluate on Test

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

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

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

```
[0]:
```

Epoch [26/50], Iter [120/209] Loss: 2.6072, average_loss: 2.8314
Epoch [26/50], Iter [125/209] Loss: 2.5009, average_loss: 2.8231
Epoch [26/50], Iter [130/209] Loss: 2.8470, average_loss: 2.8183
Epoch [26/50], Iter [135/209] Loss: 4.1249, average_loss: 2.8237
Epoch [26/50], Iter [140/209] Loss: 2.2435, average_loss: 2.8251
Epoch [26/50], Iter [145/209] Loss: 2.8255, average_loss: 2.8395
Epoch [26/50], Iter [150/209] Loss: 2.9628, average_loss: 2.8427
Epoch [26/50], Iter [155/209] Loss: 2.6527, average_loss: 2.8352
Epoch [26/50], Iter [160/209] Loss: 2.3625, average_loss: 2.8248
Epoch [26/50], Iter [165/209] Loss: 2.5859, average_loss: 2.8237
Epoch [26/50], Iter [170/209] Loss: 3.2515, average_loss: 2.8203
Epoch [26/50], Iter [175/209] Loss: 3.2258, average_loss: 2.8193
Epoch [26/50], Iter [180/209] Loss: 2.2309, average_loss: 2.8136
Epoch [26/50], Iter [185/209] Loss: 2.4934, average_loss: 2.8151
Epoch [26/50], Iter [190/209] Loss: 3.1486, average_loss: 2.8149
Epoch [26/50], Iter [195/209] Loss: 2.9744, average_loss: 2.8267
Epoch [26/50], Iter [200/209] Loss: 2.2595, average_loss: 2.8252
Epoch [26/50], Iter [205/209] Loss: 2.5911, average_loss: 2.8145
Updating best test loss: 3.58967

Starting epoch 27 / 50
Learning Rate for this epoch: 0.001
Epoch [27/50], Iter [5/209] Loss: 2.1449, average_loss: 2.2694
Epoch [27/50], Iter [10/209] Loss: 2.3921, average_loss: 2.4709
Epoch [27/50], Iter [15/209] Loss: 2.4261, average_loss: 2.5440
Epoch [27/50], Iter [20/209] Loss: 2.7300, average_loss: 2.6255
Epoch [27/50], Iter [25/209] Loss: 2.7895, average_loss: 2.5774
Epoch [27/50], Iter [30/209] Loss: 2.3790, average_loss: 2.6298
Epoch [27/50], Iter [35/209] Loss: 3.3464, average_loss: 2.6914
Epoch [27/50], Iter [40/209] Loss: 3.2473, average_loss: 2.6836
Epoch [27/50], Iter [45/209] Loss: 2.4489, average_loss: 2.6653
Epoch [27/50], Iter [50/209] Loss: 2.5968, average_loss: 2.6976
Epoch [27/50], Iter [55/209] Loss: 2.8854, average_loss: 2.7129
Epoch [27/50], Iter [60/209] Loss: 1.9494, average_loss: 2.7263
Epoch [27/50], Iter [65/209] Loss: 2.1256, average_loss: 2.7163
Epoch [27/50], Iter [70/209] Loss: 3.1236, average_loss: 2.7459
Epoch [27/50], Iter [75/209] Loss: 2.5192, average_loss: 2.7418
Epoch [27/50], Iter [80/209] Loss: 2.4851, average_loss: 2.7432
Epoch [27/50], Iter [85/209] Loss: 2.1234, average_loss: 2.7397
Epoch [27/50], Iter [90/209] Loss: 3.2360, average_loss: 2.7433
Epoch [27/50], Iter [95/209] Loss: 2.9060, average_loss: 2.7442
Epoch [27/50], Iter [100/209] Loss: 2.4869, average_loss: 2.7305
Epoch [27/50], Iter [105/209] Loss: 2.8334, average_loss: 2.7187
Epoch [27/50], Iter [110/209] Loss: 2.6731, average_loss: 2.7331
Epoch [27/50], Iter [115/209] Loss: 2.3596, average_loss: 2.7292
Epoch [27/50], Iter [120/209] Loss: 2.2931, average_loss: 2.7265
Epoch [27/50], Iter [125/209] Loss: 2.5612, average_loss: 2.7188
Epoch [27/50], Iter [130/209] Loss: 2.4275, average_loss: 2.7236
Epoch [27/50], Iter [135/209] Loss: 2.6214, average_loss: 2.7256
Epoch [27/50], Iter [140/209] Loss: 3.1026, average_loss: 2.7287
Epoch [27/50], Iter [145/209] Loss: 2.7041, average_loss: 2.7284
Epoch [27/50], Iter [150/209] Loss: 3.1201, average_loss: 2.7330
Epoch [27/50], Iter [155/209] Loss: 2.5572, average_loss: 2.7329

Epoch [27/50], Iter [160/209] Loss: 3.8827, average_loss: 2.7437
Epoch [27/50], Iter [165/209] Loss: 2.7218, average_loss: 2.7341
Epoch [27/50], Iter [170/209] Loss: 1.9510, average_loss: 2.7327
Epoch [27/50], Iter [175/209] Loss: 2.8789, average_loss: 2.7437
Epoch [27/50], Iter [180/209] Loss: 2.2491, average_loss: 2.7403
Epoch [27/50], Iter [185/209] Loss: 3.0121, average_loss: 2.7429
Epoch [27/50], Iter [190/209] Loss: 3.9350, average_loss: 2.7556
Epoch [27/50], Iter [195/209] Loss: 2.3590, average_loss: 2.7530
Epoch [27/50], Iter [200/209] Loss: 2.8494, average_loss: 2.7551
Epoch [27/50], Iter [205/209] Loss: 2.6944, average_loss: 2.7541

Starting epoch 28 / 50

Learning Rate for this epoch: 0.001

Epoch [28/50], Iter [5/209] Loss: 2.8759, average_loss: 2.8338
Epoch [28/50], Iter [10/209] Loss: 2.7573, average_loss: 2.8480
Epoch [28/50], Iter [15/209] Loss: 2.5767, average_loss: 2.7923
Epoch [28/50], Iter [20/209] Loss: 1.7888, average_loss: 2.6959
Epoch [28/50], Iter [25/209] Loss: 2.2654, average_loss: 2.7134
Epoch [28/50], Iter [30/209] Loss: 3.2452, average_loss: 2.6929
Epoch [28/50], Iter [35/209] Loss: 2.6503, average_loss: 2.6764
Epoch [28/50], Iter [40/209] Loss: 2.6758, average_loss: 2.7190
Epoch [28/50], Iter [45/209] Loss: 2.5812, average_loss: 2.7488
Epoch [28/50], Iter [50/209] Loss: 3.5458, average_loss: 2.7333
Epoch [28/50], Iter [55/209] Loss: 2.4481, average_loss: 2.7333
Epoch [28/50], Iter [60/209] Loss: 2.8355, average_loss: 2.7415
Epoch [28/50], Iter [65/209] Loss: 4.4148, average_loss: 2.7344
Epoch [28/50], Iter [70/209] Loss: 2.0673, average_loss: 2.7283
Epoch [28/50], Iter [75/209] Loss: 3.7268, average_loss: 2.7407
Epoch [28/50], Iter [80/209] Loss: 2.0068, average_loss: 2.7455
Epoch [28/50], Iter [85/209] Loss: 2.5026, average_loss: 2.7352
Epoch [28/50], Iter [90/209] Loss: 2.4011, average_loss: 2.7501
Epoch [28/50], Iter [95/209] Loss: 5.0202, average_loss: 2.7473
Epoch [28/50], Iter [100/209] Loss: 2.8742, average_loss: 2.7540
Epoch [28/50], Iter [105/209] Loss: 2.5792, average_loss: 2.7452
Epoch [28/50], Iter [110/209] Loss: 3.4698, average_loss: 2.7494
Epoch [28/50], Iter [115/209] Loss: 3.1929, average_loss: 2.7516
Epoch [28/50], Iter [120/209] Loss: 3.5037, average_loss: 2.7469
Epoch [28/50], Iter [125/209] Loss: 2.3506, average_loss: 2.7369
Epoch [28/50], Iter [130/209] Loss: 2.2369, average_loss: 2.7512
Epoch [28/50], Iter [135/209] Loss: 3.0181, average_loss: 2.7545
Epoch [28/50], Iter [140/209] Loss: 2.6501, average_loss: 2.7587
Epoch [28/50], Iter [145/209] Loss: 4.2964, average_loss: 2.7735
Epoch [28/50], Iter [150/209] Loss: 3.4279, average_loss: 2.7693
Epoch [28/50], Iter [155/209] Loss: 3.0115, average_loss: 2.7672
Epoch [28/50], Iter [160/209] Loss: 2.8702, average_loss: 2.7693
Epoch [28/50], Iter [165/209] Loss: 2.0579, average_loss: 2.7702
Epoch [28/50], Iter [170/209] Loss: 2.5874, average_loss: 2.7814
Epoch [28/50], Iter [175/209] Loss: 2.9096, average_loss: 2.7870
Epoch [28/50], Iter [180/209] Loss: 2.9470, average_loss: 2.7822
Epoch [28/50], Iter [185/209] Loss: 2.8025, average_loss: 2.7836
Epoch [28/50], Iter [190/209] Loss: 3.2402, average_loss: 2.7848
Epoch [28/50], Iter [195/209] Loss: 1.9022, average_loss: 2.7826
Epoch [28/50], Iter [200/209] Loss: 2.7553, average_loss: 2.7823

Epoch [28/50], Iter [205/209] Loss: 2.6107, average_loss: 2.7799

Starting epoch 29 / 50

Learning Rate for this epoch: 0.001

Epoch [29/50], Iter [5/209] Loss: 2.2848, average_loss: 2.3192
Epoch [29/50], Iter [10/209] Loss: 2.3792, average_loss: 2.4021
Epoch [29/50], Iter [15/209] Loss: 2.1859, average_loss: 2.6663
Epoch [29/50], Iter [20/209] Loss: 1.9241, average_loss: 2.6571
Epoch [29/50], Iter [25/209] Loss: 2.4905, average_loss: 2.6904
Epoch [29/50], Iter [30/209] Loss: 3.1985, average_loss: 2.7107
Epoch [29/50], Iter [35/209] Loss: 2.8921, average_loss: 2.7251
Epoch [29/50], Iter [40/209] Loss: 3.0923, average_loss: 2.7414
Epoch [29/50], Iter [45/209] Loss: 2.8454, average_loss: 2.7751
Epoch [29/50], Iter [50/209] Loss: 1.7773, average_loss: 2.7406
Epoch [29/50], Iter [55/209] Loss: 2.5445, average_loss: 2.7080
Epoch [29/50], Iter [60/209] Loss: 2.7932, average_loss: 2.6965
Epoch [29/50], Iter [65/209] Loss: 3.0095, average_loss: 2.6909
Epoch [29/50], Iter [70/209] Loss: 2.2810, average_loss: 2.6635
Epoch [29/50], Iter [75/209] Loss: 2.7318, average_loss: 2.6730
Epoch [29/50], Iter [80/209] Loss: 3.3089, average_loss: 2.6981
Epoch [29/50], Iter [85/209] Loss: 2.7029, average_loss: 2.7035
Epoch [29/50], Iter [90/209] Loss: 1.9595, average_loss: 2.7205
Epoch [29/50], Iter [95/209] Loss: 3.2630, average_loss: 2.7530
Epoch [29/50], Iter [100/209] Loss: 2.9943, average_loss: 2.7372
Epoch [29/50], Iter [105/209] Loss: 3.3137, average_loss: 2.7420
Epoch [29/50], Iter [110/209] Loss: 3.7364, average_loss: 2.7457
Epoch [29/50], Iter [115/209] Loss: 3.1647, average_loss: 2.7513
Epoch [29/50], Iter [120/209] Loss: 1.9528, average_loss: 2.7418
Epoch [29/50], Iter [125/209] Loss: 3.3801, average_loss: 2.7358
Epoch [29/50], Iter [130/209] Loss: 2.4106, average_loss: 2.7501
Epoch [29/50], Iter [135/209] Loss: 2.8691, average_loss: 2.7508
Epoch [29/50], Iter [140/209] Loss: 2.8005, average_loss: 2.7574
Epoch [29/50], Iter [145/209] Loss: 1.9859, average_loss: 2.7489
Epoch [29/50], Iter [150/209] Loss: 2.9519, average_loss: 2.7496
Epoch [29/50], Iter [155/209] Loss: 2.6823, average_loss: 2.7542
Epoch [29/50], Iter [160/209] Loss: 2.3704, average_loss: 2.7492
Epoch [29/50], Iter [165/209] Loss: 2.5976, average_loss: 2.7530
Epoch [29/50], Iter [170/209] Loss: 2.9783, average_loss: 2.7544
Epoch [29/50], Iter [175/209] Loss: 2.7619, average_loss: 2.7412
Epoch [29/50], Iter [180/209] Loss: 2.6019, average_loss: 2.7377
Epoch [29/50], Iter [185/209] Loss: 2.6174, average_loss: 2.7309
Epoch [29/50], Iter [190/209] Loss: 2.8944, average_loss: 2.7287
Epoch [29/50], Iter [195/209] Loss: 2.2255, average_loss: 2.7289
Epoch [29/50], Iter [200/209] Loss: 3.3083, average_loss: 2.7289
Epoch [29/50], Iter [205/209] Loss: 2.8952, average_loss: 2.7166

Updating best test loss: 3.57607

Starting epoch 30 / 50

Learning Rate for this epoch: 0.001

Epoch [30/50], Iter [5/209] Loss: 2.9147, average_loss: 2.9325
Epoch [30/50], Iter [10/209] Loss: 1.9711, average_loss: 2.8828
Epoch [30/50], Iter [15/209] Loss: 2.6019, average_loss: 2.7864

Epoch [30/50], Iter [20/209] Loss: 2.7581, average_loss: 2.6933
Epoch [30/50], Iter [25/209] Loss: 2.0732, average_loss: 2.6227
Epoch [30/50], Iter [30/209] Loss: 3.7143, average_loss: 2.6277
Epoch [30/50], Iter [35/209] Loss: 2.3194, average_loss: 2.6052
Epoch [30/50], Iter [40/209] Loss: 1.9580, average_loss: 2.5978
Epoch [30/50], Iter [45/209] Loss: 1.7236, average_loss: 2.5774
Epoch [30/50], Iter [50/209] Loss: 2.8294, average_loss: 2.5968
Epoch [30/50], Iter [55/209] Loss: 2.0140, average_loss: 2.6111
Epoch [30/50], Iter [60/209] Loss: 2.7208, average_loss: 2.6283
Epoch [30/50], Iter [65/209] Loss: 2.8532, average_loss: 2.6446
Epoch [30/50], Iter [70/209] Loss: 2.7694, average_loss: 2.6378
Epoch [30/50], Iter [75/209] Loss: 2.6904, average_loss: 2.6314
Epoch [30/50], Iter [80/209] Loss: 2.4753, average_loss: 2.6407
Epoch [30/50], Iter [85/209] Loss: 2.3206, average_loss: 2.6607
Epoch [30/50], Iter [90/209] Loss: 1.9348, average_loss: 2.6502
Epoch [30/50], Iter [95/209] Loss: 2.4343, average_loss: 2.6605
Epoch [30/50], Iter [100/209] Loss: 2.1793, average_loss: 2.6535
Epoch [30/50], Iter [105/209] Loss: 4.1267, average_loss: 2.6719
Epoch [30/50], Iter [110/209] Loss: 2.9928, average_loss: 2.6809
Epoch [30/50], Iter [115/209] Loss: 2.5122, average_loss: 2.6772
Epoch [30/50], Iter [120/209] Loss: 2.7524, average_loss: 2.6790
Epoch [30/50], Iter [125/209] Loss: 1.5819, average_loss: 2.6766
Epoch [30/50], Iter [130/209] Loss: 3.6299, average_loss: 2.6683
Epoch [30/50], Iter [135/209] Loss: 2.4303, average_loss: 2.6770
Epoch [30/50], Iter [140/209] Loss: 2.5632, average_loss: 2.6711
Epoch [30/50], Iter [145/209] Loss: 2.8462, average_loss: 2.6822
Epoch [30/50], Iter [150/209] Loss: 2.6490, average_loss: 2.6781
Epoch [30/50], Iter [155/209] Loss: 2.7869, average_loss: 2.6731
Epoch [30/50], Iter [160/209] Loss: 2.2442, average_loss: 2.6719
Epoch [30/50], Iter [165/209] Loss: 2.8035, average_loss: 2.6706
Epoch [30/50], Iter [170/209] Loss: 2.4510, average_loss: 2.6631
Epoch [30/50], Iter [175/209] Loss: 3.5093, average_loss: 2.6774
Epoch [30/50], Iter [180/209] Loss: 1.9379, average_loss: 2.6699
Epoch [30/50], Iter [185/209] Loss: 3.3893, average_loss: 2.6701
Epoch [30/50], Iter [190/209] Loss: 2.4225, average_loss: 2.6658
Epoch [30/50], Iter [195/209] Loss: 2.6474, average_loss: 2.6649
Epoch [30/50], Iter [200/209] Loss: 3.0306, average_loss: 2.6696
Epoch [30/50], Iter [205/209] Loss: 2.9279, average_loss: 2.6701

Starting epoch 31 / 50

Learning Rate for this epoch: 0.0001

Epoch [31/50], Iter [5/209] Loss: 2.4118, average_loss: 2.6969
Epoch [31/50], Iter [10/209] Loss: 3.7375, average_loss: 2.5394
Epoch [31/50], Iter [15/209] Loss: 2.7257, average_loss: 2.5153
Epoch [31/50], Iter [20/209] Loss: 3.4672, average_loss: 2.5594
Epoch [31/50], Iter [25/209] Loss: 2.9692, average_loss: 2.6355
Epoch [31/50], Iter [30/209] Loss: 2.4215, average_loss: 2.6767
Epoch [31/50], Iter [35/209] Loss: 2.1686, average_loss: 2.6121
Epoch [31/50], Iter [40/209] Loss: 2.3740, average_loss: 2.5856
Epoch [31/50], Iter [45/209] Loss: 3.3516, average_loss: 2.6220
Epoch [31/50], Iter [50/209] Loss: 2.6904, average_loss: 2.6015
Epoch [31/50], Iter [55/209] Loss: 2.9164, average_loss: 2.5900
Epoch [31/50], Iter [60/209] Loss: 2.4007, average_loss: 2.5641

Epoch [31/50], Iter [65/209] Loss: 2.7762, average_loss: 2.5902
Epoch [31/50], Iter [70/209] Loss: 2.6474, average_loss: 2.5816
Epoch [31/50], Iter [75/209] Loss: 2.5900, average_loss: 2.6008
Epoch [31/50], Iter [80/209] Loss: 2.5178, average_loss: 2.5820
Epoch [31/50], Iter [85/209] Loss: 2.1083, average_loss: 2.5777
Epoch [31/50], Iter [90/209] Loss: 2.0373, average_loss: 2.5712
Epoch [31/50], Iter [95/209] Loss: 2.1052, average_loss: 2.5732
Epoch [31/50], Iter [100/209] Loss: 2.2133, average_loss: 2.5793
Epoch [31/50], Iter [105/209] Loss: 2.9055, average_loss: 2.5822
Epoch [31/50], Iter [110/209] Loss: 2.7188, average_loss: 2.5843
Epoch [31/50], Iter [115/209] Loss: 3.0373, average_loss: 2.5927
Epoch [31/50], Iter [120/209] Loss: 2.6044, average_loss: 2.5935
Epoch [31/50], Iter [125/209] Loss: 2.4017, average_loss: 2.5916
Epoch [31/50], Iter [130/209] Loss: 2.2423, average_loss: 2.5977
Epoch [31/50], Iter [135/209] Loss: 2.0853, average_loss: 2.5930
Epoch [31/50], Iter [140/209] Loss: 1.7092, average_loss: 2.5857
Epoch [31/50], Iter [145/209] Loss: 3.3896, average_loss: 2.5955
Epoch [31/50], Iter [150/209] Loss: 3.8901, average_loss: 2.6000
Epoch [31/50], Iter [155/209] Loss: 2.3672, average_loss: 2.5879
Epoch [31/50], Iter [160/209] Loss: 2.0309, average_loss: 2.5858
Epoch [31/50], Iter [165/209] Loss: 2.6971, average_loss: 2.5906
Epoch [31/50], Iter [170/209] Loss: 2.2712, average_loss: 2.5921
Epoch [31/50], Iter [175/209] Loss: 2.6853, average_loss: 2.5940
Epoch [31/50], Iter [180/209] Loss: 3.1142, average_loss: 2.5975
Epoch [31/50], Iter [185/209] Loss: 3.3088, average_loss: 2.5968
Epoch [31/50], Iter [190/209] Loss: 1.7392, average_loss: 2.5976
Epoch [31/50], Iter [195/209] Loss: 2.1069, average_loss: 2.6026
Epoch [31/50], Iter [200/209] Loss: 2.4935, average_loss: 2.5983
Epoch [31/50], Iter [205/209] Loss: 2.2015, average_loss: 2.5902
Updating best test loss: 3.50329

Starting epoch 32 / 50
Learning Rate for this epoch: 0.0001
Epoch [32/50], Iter [5/209] Loss: 2.6424, average_loss: 2.6585
Epoch [32/50], Iter [10/209] Loss: 1.7208, average_loss: 2.4674
Epoch [32/50], Iter [15/209] Loss: 2.1564, average_loss: 2.5129
Epoch [32/50], Iter [20/209] Loss: 1.9738, average_loss: 2.4803
Epoch [32/50], Iter [25/209] Loss: 2.0314, average_loss: 2.4748
Epoch [32/50], Iter [30/209] Loss: 1.8430, average_loss: 2.4311
Epoch [32/50], Iter [35/209] Loss: 2.3490, average_loss: 2.5608
Epoch [32/50], Iter [40/209] Loss: 1.8941, average_loss: 2.5397
Epoch [32/50], Iter [45/209] Loss: 2.1489, average_loss: 2.5171
Epoch [32/50], Iter [50/209] Loss: 2.2997, average_loss: 2.4848
Epoch [32/50], Iter [55/209] Loss: 3.3229, average_loss: 2.4947
Epoch [32/50], Iter [60/209] Loss: 2.2946, average_loss: 2.4803
Epoch [32/50], Iter [65/209] Loss: 2.2676, average_loss: 2.4955
Epoch [32/50], Iter [70/209] Loss: 2.3520, average_loss: 2.5114
Epoch [32/50], Iter [75/209] Loss: 2.0725, average_loss: 2.5204
Epoch [32/50], Iter [80/209] Loss: 3.2757, average_loss: 2.5256
Epoch [32/50], Iter [85/209] Loss: 2.2658, average_loss: 2.5157
Epoch [32/50], Iter [90/209] Loss: 3.8171, average_loss: 2.5294
Epoch [32/50], Iter [95/209] Loss: 2.4487, average_loss: 2.5168
Epoch [32/50], Iter [100/209] Loss: 2.4243, average_loss: 2.5215

Epoch [32/50], Iter [105/209] Loss: 3.4570, average_loss: 2.5589
Epoch [32/50], Iter [110/209] Loss: 2.7443, average_loss: 2.5569
Epoch [32/50], Iter [115/209] Loss: 3.7079, average_loss: 2.5689
Epoch [32/50], Iter [120/209] Loss: 1.8572, average_loss: 2.5520
Epoch [32/50], Iter [125/209] Loss: 2.1534, average_loss: 2.5507
Epoch [32/50], Iter [130/209] Loss: 2.0299, average_loss: 2.5519
Epoch [32/50], Iter [135/209] Loss: 2.2338, average_loss: 2.5463
Epoch [32/50], Iter [140/209] Loss: 2.0591, average_loss: 2.5415
Epoch [32/50], Iter [145/209] Loss: 3.5791, average_loss: 2.5502
Epoch [32/50], Iter [150/209] Loss: 1.7057, average_loss: 2.5396
Epoch [32/50], Iter [155/209] Loss: 1.8778, average_loss: 2.5294
Epoch [32/50], Iter [160/209] Loss: 2.5150, average_loss: 2.5351
Epoch [32/50], Iter [165/209] Loss: 1.9121, average_loss: 2.5214
Epoch [32/50], Iter [170/209] Loss: 2.5186, average_loss: 2.5166
Epoch [32/50], Iter [175/209] Loss: 3.7256, average_loss: 2.5278
Epoch [32/50], Iter [180/209] Loss: 2.0874, average_loss: 2.5335
Epoch [32/50], Iter [185/209] Loss: 2.6981, average_loss: 2.5277
Epoch [32/50], Iter [190/209] Loss: 2.1067, average_loss: 2.5247
Epoch [32/50], Iter [195/209] Loss: 2.3497, average_loss: 2.5200
Epoch [32/50], Iter [200/209] Loss: 2.8032, average_loss: 2.5202
Epoch [32/50], Iter [205/209] Loss: 3.1820, average_loss: 2.5266

Starting epoch 33 / 50

Learning Rate for this epoch: 0.0001

Epoch [33/50], Iter [5/209] Loss: 1.3932, average_loss: 2.4913
Epoch [33/50], Iter [10/209] Loss: 2.3405, average_loss: 2.5249
Epoch [33/50], Iter [15/209] Loss: 1.8578, average_loss: 2.4605
Epoch [33/50], Iter [20/209] Loss: 2.1600, average_loss: 2.4565
Epoch [33/50], Iter [25/209] Loss: 2.3219, average_loss: 2.4438
Epoch [33/50], Iter [30/209] Loss: 2.3866, average_loss: 2.4025
Epoch [33/50], Iter [35/209] Loss: 2.6194, average_loss: 2.3864
Epoch [33/50], Iter [40/209] Loss: 2.0922, average_loss: 2.3956
Epoch [33/50], Iter [45/209] Loss: 2.8455, average_loss: 2.4433
Epoch [33/50], Iter [50/209] Loss: 2.1954, average_loss: 2.4416
Epoch [33/50], Iter [55/209] Loss: 2.6851, average_loss: 2.4377
Epoch [33/50], Iter [60/209] Loss: 1.4791, average_loss: 2.4347
Epoch [33/50], Iter [65/209] Loss: 3.4257, average_loss: 2.4552
Epoch [33/50], Iter [70/209] Loss: 3.3045, average_loss: 2.4698
Epoch [33/50], Iter [75/209] Loss: 2.7351, average_loss: 2.4930
Epoch [33/50], Iter [80/209] Loss: 2.7549, average_loss: 2.4895
Epoch [33/50], Iter [85/209] Loss: 2.3120, average_loss: 2.4802
Epoch [33/50], Iter [90/209] Loss: 2.9387, average_loss: 2.4681
Epoch [33/50], Iter [95/209] Loss: 3.0809, average_loss: 2.4780
Epoch [33/50], Iter [100/209] Loss: 2.6797, average_loss: 2.4841
Epoch [33/50], Iter [105/209] Loss: 2.3386, average_loss: 2.4743
Epoch [33/50], Iter [110/209] Loss: 2.3276, average_loss: 2.4829
Epoch [33/50], Iter [115/209] Loss: 2.4720, average_loss: 2.4910
Epoch [33/50], Iter [120/209] Loss: 2.1289, average_loss: 2.4912
Epoch [33/50], Iter [125/209] Loss: 1.8327, average_loss: 2.4915
Epoch [33/50], Iter [130/209] Loss: 1.9732, average_loss: 2.4829
Epoch [33/50], Iter [135/209] Loss: 2.2944, average_loss: 2.4770
Epoch [33/50], Iter [140/209] Loss: 2.6445, average_loss: 2.4806
Epoch [33/50], Iter [145/209] Loss: 2.5886, average_loss: 2.4810

Epoch [33/50], Iter [150/209] Loss: 2.1145, average_loss: 2.4753
Epoch [33/50], Iter [155/209] Loss: 2.1784, average_loss: 2.4641
Epoch [33/50], Iter [160/209] Loss: 3.4071, average_loss: 2.4759
Epoch [33/50], Iter [165/209] Loss: 3.1549, average_loss: 2.4859
Epoch [33/50], Iter [170/209] Loss: 2.6510, average_loss: 2.4877
Epoch [33/50], Iter [175/209] Loss: 2.2358, average_loss: 2.4937
Epoch [33/50], Iter [180/209] Loss: 2.3677, average_loss: 2.4935
Epoch [33/50], Iter [185/209] Loss: 2.9839, average_loss: 2.5002
Epoch [33/50], Iter [190/209] Loss: 1.5660, average_loss: 2.4997
Epoch [33/50], Iter [195/209] Loss: 2.6683, average_loss: 2.5005
Epoch [33/50], Iter [200/209] Loss: 2.2646, average_loss: 2.5002
Epoch [33/50], Iter [205/209] Loss: 1.9136, average_loss: 2.4978
Updating best test loss: 3.50014

Starting epoch 34 / 50
Learning Rate for this epoch: 0.0001
Epoch [34/50], Iter [5/209] Loss: 3.1191, average_loss: 2.4205
Epoch [34/50], Iter [10/209] Loss: 2.5433, average_loss: 2.4067
Epoch [34/50], Iter [15/209] Loss: 2.2924, average_loss: 2.4335
Epoch [34/50], Iter [20/209] Loss: 2.3910, average_loss: 2.4305
Epoch [34/50], Iter [25/209] Loss: 2.2530, average_loss: 2.5438
Epoch [34/50], Iter [30/209] Loss: 4.0513, average_loss: 2.6336
Epoch [34/50], Iter [35/209] Loss: 2.3213, average_loss: 2.5984
Epoch [34/50], Iter [40/209] Loss: 2.5166, average_loss: 2.6065
Epoch [34/50], Iter [45/209] Loss: 2.1089, average_loss: 2.5903
Epoch [34/50], Iter [50/209] Loss: 2.7645, average_loss: 2.5801
Epoch [34/50], Iter [55/209] Loss: 2.0238, average_loss: 2.5568
Epoch [34/50], Iter [60/209] Loss: 2.4439, average_loss: 2.5579
Epoch [34/50], Iter [65/209] Loss: 3.2082, average_loss: 2.5646
Epoch [34/50], Iter [70/209] Loss: 2.1574, average_loss: 2.5489
Epoch [34/50], Iter [75/209] Loss: 2.6361, average_loss: 2.5369
Epoch [34/50], Iter [80/209] Loss: 2.5593, average_loss: 2.5323
Epoch [34/50], Iter [85/209] Loss: 2.8896, average_loss: 2.5433
Epoch [34/50], Iter [90/209] Loss: 3.1362, average_loss: 2.5307
Epoch [34/50], Iter [95/209] Loss: 2.2380, average_loss: 2.5284
Epoch [34/50], Iter [100/209] Loss: 2.0623, average_loss: 2.5064
Epoch [34/50], Iter [105/209] Loss: 2.4031, average_loss: 2.5039
Epoch [34/50], Iter [110/209] Loss: 2.3390, average_loss: 2.5081
Epoch [34/50], Iter [115/209] Loss: 2.6554, average_loss: 2.5165
Epoch [34/50], Iter [120/209] Loss: 2.2327, average_loss: 2.5052
Epoch [34/50], Iter [125/209] Loss: 2.1281, average_loss: 2.4969
Epoch [34/50], Iter [130/209] Loss: 3.1841, average_loss: 2.4921
Epoch [34/50], Iter [135/209] Loss: 2.4986, average_loss: 2.5012
Epoch [34/50], Iter [140/209] Loss: 2.9481, average_loss: 2.4944
Epoch [34/50], Iter [145/209] Loss: 2.4088, average_loss: 2.4870
Epoch [34/50], Iter [150/209] Loss: 2.5510, average_loss: 2.4999
Epoch [34/50], Iter [155/209] Loss: 3.3082, average_loss: 2.5032
Epoch [34/50], Iter [160/209] Loss: 2.3095, average_loss: 2.4971
Epoch [34/50], Iter [165/209] Loss: 2.2102, average_loss: 2.4982
Epoch [34/50], Iter [170/209] Loss: 3.5873, average_loss: 2.5137
Epoch [34/50], Iter [175/209] Loss: 1.9432, average_loss: 2.5117
Epoch [34/50], Iter [180/209] Loss: 2.5601, average_loss: 2.5156
Epoch [34/50], Iter [185/209] Loss: 2.2666, average_loss: 2.5097

Epoch [34/50], Iter [190/209] Loss: 2.1152, average_loss: 2.5123
Epoch [34/50], Iter [195/209] Loss: 2.4283, average_loss: 2.5027
Epoch [34/50], Iter [200/209] Loss: 1.9664, average_loss: 2.4948
Epoch [34/50], Iter [205/209] Loss: 2.3847, average_loss: 2.4974

Starting epoch 35 / 50
Learning Rate for this epoch: 0.0001
Epoch [35/50], Iter [5/209] Loss: 1.9913, average_loss: 2.2531
Epoch [35/50], Iter [10/209] Loss: 2.7466, average_loss: 2.3273
Epoch [35/50], Iter [15/209] Loss: 2.5058, average_loss: 2.4170
Epoch [35/50], Iter [20/209] Loss: 1.8350, average_loss: 2.3913
Epoch [35/50], Iter [25/209] Loss: 2.5554, average_loss: 2.3957
Epoch [35/50], Iter [30/209] Loss: 2.0526, average_loss: 2.3962
Epoch [35/50], Iter [35/209] Loss: 2.7118, average_loss: 2.4325
Epoch [35/50], Iter [40/209] Loss: 2.1628, average_loss: 2.4196
Epoch [35/50], Iter [45/209] Loss: 2.0986, average_loss: 2.4191
Epoch [35/50], Iter [50/209] Loss: 2.0307, average_loss: 2.4276
Epoch [35/50], Iter [55/209] Loss: 2.4122, average_loss: 2.4915
Epoch [35/50], Iter [60/209] Loss: 2.6195, average_loss: 2.4754
Epoch [35/50], Iter [65/209] Loss: 2.5377, average_loss: 2.4858
Epoch [35/50], Iter [70/209] Loss: 2.0693, average_loss: 2.4774
Epoch [35/50], Iter [75/209] Loss: 2.3653, average_loss: 2.4735
Epoch [35/50], Iter [80/209] Loss: 2.1711, average_loss: 2.4700
Epoch [35/50], Iter [85/209] Loss: 2.8213, average_loss: 2.4517
Epoch [35/50], Iter [90/209] Loss: 2.5542, average_loss: 2.4549
Epoch [35/50], Iter [95/209] Loss: 3.0460, average_loss: 2.4587
Epoch [35/50], Iter [100/209] Loss: 2.4361, average_loss: 2.4492
Epoch [35/50], Iter [105/209] Loss: 2.5146, average_loss: 2.4585
Epoch [35/50], Iter [110/209] Loss: 3.1104, average_loss: 2.4775
Epoch [35/50], Iter [115/209] Loss: 2.8061, average_loss: 2.4827
Epoch [35/50], Iter [120/209] Loss: 2.2799, average_loss: 2.4822
Epoch [35/50], Iter [125/209] Loss: 2.9422, average_loss: 2.4784
Epoch [35/50], Iter [130/209] Loss: 2.0814, average_loss: 2.4641
Epoch [35/50], Iter [135/209] Loss: 2.4610, average_loss: 2.4740
Epoch [35/50], Iter [140/209] Loss: 2.5429, average_loss: 2.4786
Epoch [35/50], Iter [145/209] Loss: 4.0386, average_loss: 2.4788
Epoch [35/50], Iter [150/209] Loss: 1.9617, average_loss: 2.4877
Epoch [35/50], Iter [155/209] Loss: 2.0431, average_loss: 2.4848
Epoch [35/50], Iter [160/209] Loss: 1.8321, average_loss: 2.4823
Epoch [35/50], Iter [165/209] Loss: 2.4212, average_loss: 2.4809
Epoch [35/50], Iter [170/209] Loss: 2.3909, average_loss: 2.4931
Epoch [35/50], Iter [175/209] Loss: 2.3171, average_loss: 2.4980
Epoch [35/50], Iter [180/209] Loss: 2.2329, average_loss: 2.4943
Epoch [35/50], Iter [185/209] Loss: 1.7616, average_loss: 2.4927
Epoch [35/50], Iter [190/209] Loss: 4.1414, average_loss: 2.4967
Epoch [35/50], Iter [195/209] Loss: 2.6929, average_loss: 2.4966
Epoch [35/50], Iter [200/209] Loss: 2.4299, average_loss: 2.5020
Epoch [35/50], Iter [205/209] Loss: 2.2783, average_loss: 2.5043

Starting epoch 36 / 50
Learning Rate for this epoch: 0.0001
Epoch [36/50], Iter [5/209] Loss: 1.8453, average_loss: 2.1883

Epoch [36/50], Iter [10/209] Loss: 3.3565, average_loss: 2.2727
Epoch [36/50], Iter [15/209] Loss: 2.8050, average_loss: 2.3381
Epoch [36/50], Iter [20/209] Loss: 2.0339, average_loss: 2.3157
Epoch [36/50], Iter [25/209] Loss: 2.2788, average_loss: 2.3545
Epoch [36/50], Iter [30/209] Loss: 2.2140, average_loss: 2.3283
Epoch [36/50], Iter [35/209] Loss: 1.8269, average_loss: 2.2983
Epoch [36/50], Iter [40/209] Loss: 1.9170, average_loss: 2.3133
Epoch [36/50], Iter [45/209] Loss: 2.0598, average_loss: 2.3407
Epoch [36/50], Iter [50/209] Loss: 2.5396, average_loss: 2.3414
Epoch [36/50], Iter [55/209] Loss: 3.3617, average_loss: 2.3700
Epoch [36/50], Iter [60/209] Loss: 3.1808, average_loss: 2.3971
Epoch [36/50], Iter [65/209] Loss: 1.8743, average_loss: 2.3828
Epoch [36/50], Iter [70/209] Loss: 1.8627, average_loss: 2.3672
Epoch [36/50], Iter [75/209] Loss: 3.0069, average_loss: 2.3824
Epoch [36/50], Iter [80/209] Loss: 1.9413, average_loss: 2.3775
Epoch [36/50], Iter [85/209] Loss: 2.2428, average_loss: 2.3594
Epoch [36/50], Iter [90/209] Loss: 3.2804, average_loss: 2.3712
Epoch [36/50], Iter [95/209] Loss: 2.0584, average_loss: 2.3679
Epoch [36/50], Iter [100/209] Loss: 2.4308, average_loss: 2.3723
Epoch [36/50], Iter [105/209] Loss: 2.6177, average_loss: 2.3694
Epoch [36/50], Iter [110/209] Loss: 4.0451, average_loss: 2.3958
Epoch [36/50], Iter [115/209] Loss: 1.9705, average_loss: 2.3954
Epoch [36/50], Iter [120/209] Loss: 2.5464, average_loss: 2.4185
Epoch [36/50], Iter [125/209] Loss: 2.9036, average_loss: 2.4313
Epoch [36/50], Iter [130/209] Loss: 2.4966, average_loss: 2.4265
Epoch [36/50], Iter [135/209] Loss: 2.6057, average_loss: 2.4312
Epoch [36/50], Iter [140/209] Loss: 2.7013, average_loss: 2.4257
Epoch [36/50], Iter [145/209] Loss: 2.5815, average_loss: 2.4330
Epoch [36/50], Iter [150/209] Loss: 2.8338, average_loss: 2.4365
Epoch [36/50], Iter [155/209] Loss: 2.1877, average_loss: 2.4393
Epoch [36/50], Iter [160/209] Loss: 2.2813, average_loss: 2.4421
Epoch [36/50], Iter [165/209] Loss: 2.4978, average_loss: 2.4436
Epoch [36/50], Iter [170/209] Loss: 2.4916, average_loss: 2.4458
Epoch [36/50], Iter [175/209] Loss: 3.1240, average_loss: 2.4562
Epoch [36/50], Iter [180/209] Loss: 2.1804, average_loss: 2.4525
Epoch [36/50], Iter [185/209] Loss: 2.3829, average_loss: 2.4619
Epoch [36/50], Iter [190/209] Loss: 2.3767, average_loss: 2.4605
Epoch [36/50], Iter [195/209] Loss: 2.3352, average_loss: 2.4630
Epoch [36/50], Iter [200/209] Loss: 3.2247, average_loss: 2.4641
Epoch [36/50], Iter [205/209] Loss: 2.9709, average_loss: 2.4672

Starting epoch 37 / 50

Learning Rate for this epoch: 0.0001

Epoch [37/50], Iter [5/209] Loss: 2.6979, average_loss: 2.5359
Epoch [37/50], Iter [10/209] Loss: 2.3341, average_loss: 2.4676
Epoch [37/50], Iter [15/209] Loss: 2.2390, average_loss: 2.3909
Epoch [37/50], Iter [20/209] Loss: 2.7240, average_loss: 2.3644
Epoch [37/50], Iter [25/209] Loss: 2.4225, average_loss: 2.3530
Epoch [37/50], Iter [30/209] Loss: 2.1308, average_loss: 2.3895
Epoch [37/50], Iter [35/209] Loss: 2.5067, average_loss: 2.4529
Epoch [37/50], Iter [40/209] Loss: 2.9478, average_loss: 2.4263
Epoch [37/50], Iter [45/209] Loss: 2.4612, average_loss: 2.4499
Epoch [37/50], Iter [50/209] Loss: 2.3471, average_loss: 2.4413

Epoch [37/50], Iter [55/209] Loss: 1.5769, average_loss: 2.4427
Epoch [37/50], Iter [60/209] Loss: 2.3068, average_loss: 2.4160
Epoch [37/50], Iter [65/209] Loss: 2.5598, average_loss: 2.4378
Epoch [37/50], Iter [70/209] Loss: 2.3809, average_loss: 2.4451
Epoch [37/50], Iter [75/209] Loss: 2.6895, average_loss: 2.4281
Epoch [37/50], Iter [80/209] Loss: 1.9971, average_loss: 2.4081
Epoch [37/50], Iter [85/209] Loss: 2.4322, average_loss: 2.4125
Epoch [37/50], Iter [90/209] Loss: 2.3008, average_loss: 2.4156
Epoch [37/50], Iter [95/209] Loss: 2.1526, average_loss: 2.4286
Epoch [37/50], Iter [100/209] Loss: 2.3208, average_loss: 2.4330
Epoch [37/50], Iter [105/209] Loss: 2.9324, average_loss: 2.4246
Epoch [37/50], Iter [110/209] Loss: 2.3383, average_loss: 2.4364
Epoch [37/50], Iter [115/209] Loss: 2.6370, average_loss: 2.4261
Epoch [37/50], Iter [120/209] Loss: 2.7671, average_loss: 2.4283
Epoch [37/50], Iter [125/209] Loss: 2.5604, average_loss: 2.4392
Epoch [37/50], Iter [130/209] Loss: 2.4484, average_loss: 2.4529
Epoch [37/50], Iter [135/209] Loss: 2.7386, average_loss: 2.4477
Epoch [37/50], Iter [140/209] Loss: 2.4631, average_loss: 2.4583
Epoch [37/50], Iter [145/209] Loss: 3.1033, average_loss: 2.4594
Epoch [37/50], Iter [150/209] Loss: 2.7684, average_loss: 2.4580
Epoch [37/50], Iter [155/209] Loss: 1.9351, average_loss: 2.4524
Epoch [37/50], Iter [160/209] Loss: 2.7935, average_loss: 2.4600
Epoch [37/50], Iter [165/209] Loss: 2.1295, average_loss: 2.4546
Epoch [37/50], Iter [170/209] Loss: 2.4653, average_loss: 2.4514
Epoch [37/50], Iter [175/209] Loss: 1.7348, average_loss: 2.4433
Epoch [37/50], Iter [180/209] Loss: 2.6875, average_loss: 2.4483
Epoch [37/50], Iter [185/209] Loss: 2.4526, average_loss: 2.4552
Epoch [37/50], Iter [190/209] Loss: 1.7973, average_loss: 2.4505
Epoch [37/50], Iter [195/209] Loss: 1.9824, average_loss: 2.4489
Epoch [37/50], Iter [200/209] Loss: 2.1637, average_loss: 2.4495
Epoch [37/50], Iter [205/209] Loss: 2.5528, average_loss: 2.4470
Updating best test loss: 3.48114

Starting epoch 38 / 50
Learning Rate for this epoch: 0.0001
Epoch [38/50], Iter [5/209] Loss: 2.6368, average_loss: 2.4672
Epoch [38/50], Iter [10/209] Loss: 3.3352, average_loss: 2.6071
Epoch [38/50], Iter [15/209] Loss: 2.4362, average_loss: 2.4873
Epoch [38/50], Iter [20/209] Loss: 2.7113, average_loss: 2.4525
Epoch [38/50], Iter [25/209] Loss: 2.3923, average_loss: 2.4063
Epoch [38/50], Iter [30/209] Loss: 2.1844, average_loss: 2.3975
Epoch [38/50], Iter [35/209] Loss: 2.7817, average_loss: 2.3992
Epoch [38/50], Iter [40/209] Loss: 2.1456, average_loss: 2.3987
Epoch [38/50], Iter [45/209] Loss: 2.7135, average_loss: 2.3778
Epoch [38/50], Iter [50/209] Loss: 2.2747, average_loss: 2.3998
Epoch [38/50], Iter [55/209] Loss: 2.7969, average_loss: 2.4570
Epoch [38/50], Iter [60/209] Loss: 1.9977, average_loss: 2.4425
Epoch [38/50], Iter [65/209] Loss: 2.0945, average_loss: 2.4410
Epoch [38/50], Iter [70/209] Loss: 2.5191, average_loss: 2.4586
Epoch [38/50], Iter [75/209] Loss: 3.5421, average_loss: 2.4805
Epoch [38/50], Iter [80/209] Loss: 3.0068, average_loss: 2.4918
Epoch [38/50], Iter [85/209] Loss: 2.6377, average_loss: 2.4955
Epoch [38/50], Iter [90/209] Loss: 2.8673, average_loss: 2.5145

Epoch [38/50], Iter [95/209] Loss: 2.4890, average_loss: 2.5047
Epoch [38/50], Iter [100/209] Loss: 2.3871, average_loss: 2.4857
Epoch [38/50], Iter [105/209] Loss: 1.9691, average_loss: 2.4869
Epoch [38/50], Iter [110/209] Loss: 2.3350, average_loss: 2.4918
Epoch [38/50], Iter [115/209] Loss: 2.3408, average_loss: 2.4897
Epoch [38/50], Iter [120/209] Loss: 2.1623, average_loss: 2.4935
Epoch [38/50], Iter [125/209] Loss: 2.9656, average_loss: 2.4964
Epoch [38/50], Iter [130/209] Loss: 2.3730, average_loss: 2.4904
Epoch [38/50], Iter [135/209] Loss: 2.9415, average_loss: 2.4988
Epoch [38/50], Iter [140/209] Loss: 2.8487, average_loss: 2.4997
Epoch [38/50], Iter [145/209] Loss: 1.3594, average_loss: 2.4889
Epoch [38/50], Iter [150/209] Loss: 2.2481, average_loss: 2.4855
Epoch [38/50], Iter [155/209] Loss: 2.7956, average_loss: 2.4873
Epoch [38/50], Iter [160/209] Loss: 2.0201, average_loss: 2.4907
Epoch [38/50], Iter [165/209] Loss: 2.1781, average_loss: 2.4794
Epoch [38/50], Iter [170/209] Loss: 2.0824, average_loss: 2.4740
Epoch [38/50], Iter [175/209] Loss: 2.5720, average_loss: 2.4738
Epoch [38/50], Iter [180/209] Loss: 2.5699, average_loss: 2.4675
Epoch [38/50], Iter [185/209] Loss: 2.4912, average_loss: 2.4652
Epoch [38/50], Iter [190/209] Loss: 2.3928, average_loss: 2.4639
Epoch [38/50], Iter [195/209] Loss: 1.7917, average_loss: 2.4578
Epoch [38/50], Iter [200/209] Loss: 1.9371, average_loss: 2.4524
Epoch [38/50], Iter [205/209] Loss: 2.2416, average_loss: 2.4576

Starting epoch 39 / 50

Learning Rate for this epoch: 0.0001

Epoch [39/50], Iter [5/209] Loss: 2.0360, average_loss: 2.7350
Epoch [39/50], Iter [10/209] Loss: 2.2292, average_loss: 2.6948
Epoch [39/50], Iter [15/209] Loss: 1.9681, average_loss: 2.6099
Epoch [39/50], Iter [20/209] Loss: 2.8435, average_loss: 2.5740
Epoch [39/50], Iter [25/209] Loss: 3.7128, average_loss: 2.5911
Epoch [39/50], Iter [30/209] Loss: 2.6678, average_loss: 2.6513
Epoch [39/50], Iter [35/209] Loss: 2.1968, average_loss: 2.5956
Epoch [39/50], Iter [40/209] Loss: 2.2724, average_loss: 2.5462
Epoch [39/50], Iter [45/209] Loss: 2.0085, average_loss: 2.5886
Epoch [39/50], Iter [50/209] Loss: 2.0100, average_loss: 2.5614
Epoch [39/50], Iter [55/209] Loss: 2.0572, average_loss: 2.5488
Epoch [39/50], Iter [60/209] Loss: 2.9599, average_loss: 2.5389
Epoch [39/50], Iter [65/209] Loss: 2.5225, average_loss: 2.5556
Epoch [39/50], Iter [70/209] Loss: 1.7802, average_loss: 2.5125
Epoch [39/50], Iter [75/209] Loss: 2.3716, average_loss: 2.5194
Epoch [39/50], Iter [80/209] Loss: 2.3501, average_loss: 2.4954
Epoch [39/50], Iter [85/209] Loss: 2.5011, average_loss: 2.4950
Epoch [39/50], Iter [90/209] Loss: 1.7599, average_loss: 2.4917
Epoch [39/50], Iter [95/209] Loss: 3.0961, average_loss: 2.4974
Epoch [39/50], Iter [100/209] Loss: 2.2816, average_loss: 2.4940
Epoch [39/50], Iter [105/209] Loss: 2.1855, average_loss: 2.4982
Epoch [39/50], Iter [110/209] Loss: 2.8678, average_loss: 2.5013
Epoch [39/50], Iter [115/209] Loss: 2.0607, average_loss: 2.5013
Epoch [39/50], Iter [120/209] Loss: 1.7983, average_loss: 2.4898
Epoch [39/50], Iter [125/209] Loss: 2.7440, average_loss: 2.4818
Epoch [39/50], Iter [130/209] Loss: 1.8471, average_loss: 2.4708
Epoch [39/50], Iter [135/209] Loss: 2.9421, average_loss: 2.4733

Epoch [39/50], Iter [140/209] Loss: 3.0921, average_loss: 2.4794
Epoch [39/50], Iter [145/209] Loss: 2.1943, average_loss: 2.4813
Epoch [39/50], Iter [150/209] Loss: 2.1721, average_loss: 2.4901
Epoch [39/50], Iter [155/209] Loss: 2.0647, average_loss: 2.4891
Epoch [39/50], Iter [160/209] Loss: 1.8625, average_loss: 2.4874
Epoch [39/50], Iter [165/209] Loss: 2.8421, average_loss: 2.4870
Epoch [39/50], Iter [170/209] Loss: 2.2623, average_loss: 2.4742
Epoch [39/50], Iter [175/209] Loss: 2.0542, average_loss: 2.4665
Epoch [39/50], Iter [180/209] Loss: 2.7663, average_loss: 2.4703
Epoch [39/50], Iter [185/209] Loss: 2.8120, average_loss: 2.4663
Epoch [39/50], Iter [190/209] Loss: 1.6185, average_loss: 2.4573
Epoch [39/50], Iter [195/209] Loss: 2.8142, average_loss: 2.4674
Epoch [39/50], Iter [200/209] Loss: 1.9923, average_loss: 2.4705
Epoch [39/50], Iter [205/209] Loss: 4.1128, average_loss: 2.4696

Starting epoch 40 / 50

Learning Rate for this epoch: 0.0001

Epoch [40/50], Iter [5/209] Loss: 1.7832, average_loss: 2.2520
Epoch [40/50], Iter [10/209] Loss: 2.2599, average_loss: 2.3011
Epoch [40/50], Iter [15/209] Loss: 2.7733, average_loss: 2.2953
Epoch [40/50], Iter [20/209] Loss: 2.3280, average_loss: 2.2474
Epoch [40/50], Iter [25/209] Loss: 2.2334, average_loss: 2.2572
Epoch [40/50], Iter [30/209] Loss: 1.7238, average_loss: 2.2876
Epoch [40/50], Iter [35/209] Loss: 2.4984, average_loss: 2.2674
Epoch [40/50], Iter [40/209] Loss: 2.2665, average_loss: 2.2936
Epoch [40/50], Iter [45/209] Loss: 2.3439, average_loss: 2.3191
Epoch [40/50], Iter [50/209] Loss: 2.5259, average_loss: 2.3153
Epoch [40/50], Iter [55/209] Loss: 2.3396, average_loss: 2.3447
Epoch [40/50], Iter [60/209] Loss: 2.2926, average_loss: 2.3415
Epoch [40/50], Iter [65/209] Loss: 2.3710, average_loss: 2.3619
Epoch [40/50], Iter [70/209] Loss: 2.2106, average_loss: 2.3461
Epoch [40/50], Iter [75/209] Loss: 2.2157, average_loss: 2.3547
Epoch [40/50], Iter [80/209] Loss: 3.5581, average_loss: 2.3691
Epoch [40/50], Iter [85/209] Loss: 2.2424, average_loss: 2.3786
Epoch [40/50], Iter [90/209] Loss: 2.1330, average_loss: 2.3828
Epoch [40/50], Iter [95/209] Loss: 2.2399, average_loss: 2.3971
Epoch [40/50], Iter [100/209] Loss: 3.4931, average_loss: 2.4061
Epoch [40/50], Iter [105/209] Loss: 2.1936, average_loss: 2.4016
Epoch [40/50], Iter [110/209] Loss: 2.9290, average_loss: 2.4135
Epoch [40/50], Iter [115/209] Loss: 2.4902, average_loss: 2.4056
Epoch [40/50], Iter [120/209] Loss: 3.0180, average_loss: 2.4137
Epoch [40/50], Iter [125/209] Loss: 2.9553, average_loss: 2.4091
Epoch [40/50], Iter [130/209] Loss: 2.0653, average_loss: 2.4095
Epoch [40/50], Iter [135/209] Loss: 2.7131, average_loss: 2.4076
Epoch [40/50], Iter [140/209] Loss: 3.2080, average_loss: 2.4099
Epoch [40/50], Iter [145/209] Loss: 2.1610, average_loss: 2.4125
Epoch [40/50], Iter [150/209] Loss: 1.6901, average_loss: 2.4012
Epoch [40/50], Iter [155/209] Loss: 1.8615, average_loss: 2.4042
Epoch [40/50], Iter [160/209] Loss: 2.0815, average_loss: 2.3964
Epoch [40/50], Iter [165/209] Loss: 2.9743, average_loss: 2.3967
Epoch [40/50], Iter [170/209] Loss: 1.9850, average_loss: 2.4068
Epoch [40/50], Iter [175/209] Loss: 2.0273, average_loss: 2.4062
Epoch [40/50], Iter [180/209] Loss: 3.0499, average_loss: 2.4065

Epoch [40/50], Iter [185/209] Loss: 1.9471, average_loss: 2.4073
Epoch [40/50], Iter [190/209] Loss: 2.7077, average_loss: 2.4161
Epoch [40/50], Iter [195/209] Loss: 3.2467, average_loss: 2.4152
Epoch [40/50], Iter [200/209] Loss: 2.8179, average_loss: 2.4255
Epoch [40/50], Iter [205/209] Loss: 3.0118, average_loss: 2.4321

Starting epoch 41 / 50

Learning Rate for this epoch: 1e-05

Epoch [41/50], Iter [5/209] Loss: 1.7591, average_loss: 2.0426
Epoch [41/50], Iter [10/209] Loss: 1.7714, average_loss: 2.0536
Epoch [41/50], Iter [15/209] Loss: 2.5716, average_loss: 2.2012
Epoch [41/50], Iter [20/209] Loss: 2.7389, average_loss: 2.2715
Epoch [41/50], Iter [25/209] Loss: 3.0016, average_loss: 2.3245
Epoch [41/50], Iter [30/209] Loss: 3.0203, average_loss: 2.4280
Epoch [41/50], Iter [35/209] Loss: 2.4911, average_loss: 2.4099
Epoch [41/50], Iter [40/209] Loss: 3.2670, average_loss: 2.4137
Epoch [41/50], Iter [45/209] Loss: 2.6139, average_loss: 2.4453
Epoch [41/50], Iter [50/209] Loss: 2.2330, average_loss: 2.4223
Epoch [41/50], Iter [55/209] Loss: 2.2667, average_loss: 2.4093
Epoch [41/50], Iter [60/209] Loss: 2.8259, average_loss: 2.4006
Epoch [41/50], Iter [65/209] Loss: 1.7359, average_loss: 2.4087
Epoch [41/50], Iter [70/209] Loss: 2.4322, average_loss: 2.4083
Epoch [41/50], Iter [75/209] Loss: 3.0464, average_loss: 2.4469
Epoch [41/50], Iter [80/209] Loss: 2.4361, average_loss: 2.4340
Epoch [41/50], Iter [85/209] Loss: 2.3656, average_loss: 2.4275
Epoch [41/50], Iter [90/209] Loss: 1.8171, average_loss: 2.4263
Epoch [41/50], Iter [95/209] Loss: 2.0202, average_loss: 2.4312
Epoch [41/50], Iter [100/209] Loss: 2.3223, average_loss: 2.4200
Epoch [41/50], Iter [105/209] Loss: 2.1423, average_loss: 2.4072
Epoch [41/50], Iter [110/209] Loss: 1.9555, average_loss: 2.4016
Epoch [41/50], Iter [115/209] Loss: 2.0696, average_loss: 2.3932
Epoch [41/50], Iter [120/209] Loss: 2.3299, average_loss: 2.3903
Epoch [41/50], Iter [125/209] Loss: 2.6255, average_loss: 2.3999
Epoch [41/50], Iter [130/209] Loss: 2.5665, average_loss: 2.4043
Epoch [41/50], Iter [135/209] Loss: 2.3460, average_loss: 2.3990
Epoch [41/50], Iter [140/209] Loss: 2.7273, average_loss: 2.3990
Epoch [41/50], Iter [145/209] Loss: 1.9954, average_loss: 2.4005
Epoch [41/50], Iter [150/209] Loss: 2.0183, average_loss: 2.4023
Epoch [41/50], Iter [155/209] Loss: 2.9077, average_loss: 2.3923
Epoch [41/50], Iter [160/209] Loss: 1.9582, average_loss: 2.3927
Epoch [41/50], Iter [165/209] Loss: 1.8578, average_loss: 2.3876
Epoch [41/50], Iter [170/209] Loss: 2.9013, average_loss: 2.3928
Epoch [41/50], Iter [175/209] Loss: 2.4955, average_loss: 2.4048
Epoch [41/50], Iter [180/209] Loss: 1.8205, average_loss: 2.4087
Epoch [41/50], Iter [185/209] Loss: 2.0985, average_loss: 2.4133
Epoch [41/50], Iter [190/209] Loss: 2.0567, average_loss: 2.4121
Epoch [41/50], Iter [195/209] Loss: 2.5733, average_loss: 2.4076
Epoch [41/50], Iter [200/209] Loss: 2.5071, average_loss: 2.3992
Epoch [41/50], Iter [205/209] Loss: 3.4215, average_loss: 2.4010

Starting epoch 42 / 50

Learning Rate for this epoch: 1e-05

Epoch [42/50], Iter [5/209] Loss: 2.5811, average_loss: 2.4573
Epoch [42/50], Iter [10/209] Loss: 2.7971, average_loss: 2.5907
Epoch [42/50], Iter [15/209] Loss: 2.1957, average_loss: 2.4811
Epoch [42/50], Iter [20/209] Loss: 1.9170, average_loss: 2.3795
Epoch [42/50], Iter [25/209] Loss: 2.3971, average_loss: 2.3223
Epoch [42/50], Iter [30/209] Loss: 2.1520, average_loss: 2.3860
Epoch [42/50], Iter [35/209] Loss: 2.7025, average_loss: 2.3656
Epoch [42/50], Iter [40/209] Loss: 1.9472, average_loss: 2.3310
Epoch [42/50], Iter [45/209] Loss: 2.2332, average_loss: 2.3729
Epoch [42/50], Iter [50/209] Loss: 3.0751, average_loss: 2.4000
Epoch [42/50], Iter [55/209] Loss: 2.2538, average_loss: 2.3980
Epoch [42/50], Iter [60/209] Loss: 2.3906, average_loss: 2.3894
Epoch [42/50], Iter [65/209] Loss: 2.3977, average_loss: 2.3854
Epoch [42/50], Iter [70/209] Loss: 2.1687, average_loss: 2.3963
Epoch [42/50], Iter [75/209] Loss: 1.7657, average_loss: 2.4078
Epoch [42/50], Iter [80/209] Loss: 2.1603, average_loss: 2.3965
Epoch [42/50], Iter [85/209] Loss: 2.6718, average_loss: 2.3926
Epoch [42/50], Iter [90/209] Loss: 3.5454, average_loss: 2.4027
Epoch [42/50], Iter [95/209] Loss: 2.3090, average_loss: 2.3959
Epoch [42/50], Iter [100/209] Loss: 1.9575, average_loss: 2.3886
Epoch [42/50], Iter [105/209] Loss: 2.8082, average_loss: 2.3944
Epoch [42/50], Iter [110/209] Loss: 2.1404, average_loss: 2.3953
Epoch [42/50], Iter [115/209] Loss: 2.1051, average_loss: 2.3903
Epoch [42/50], Iter [120/209] Loss: 2.3302, average_loss: 2.3999
Epoch [42/50], Iter [125/209] Loss: 2.6775, average_loss: 2.3960
Epoch [42/50], Iter [130/209] Loss: 2.6142, average_loss: 2.4015
Epoch [42/50], Iter [135/209] Loss: 2.9398, average_loss: 2.3953
Epoch [42/50], Iter [140/209] Loss: 2.5678, average_loss: 2.3998
Epoch [42/50], Iter [145/209] Loss: 4.3311, average_loss: 2.4039
Epoch [42/50], Iter [150/209] Loss: 2.1558, average_loss: 2.4107
Epoch [42/50], Iter [155/209] Loss: 2.3154, average_loss: 2.4076
Epoch [42/50], Iter [160/209] Loss: 3.8717, average_loss: 2.4192
Epoch [42/50], Iter [165/209] Loss: 2.0262, average_loss: 2.4223
Epoch [42/50], Iter [170/209] Loss: 2.7620, average_loss: 2.4299
Epoch [42/50], Iter [175/209] Loss: 2.0985, average_loss: 2.4374
Epoch [42/50], Iter [180/209] Loss: 2.8051, average_loss: 2.4379
Epoch [42/50], Iter [185/209] Loss: 2.4081, average_loss: 2.4360
Epoch [42/50], Iter [190/209] Loss: 2.3480, average_loss: 2.4257
Epoch [42/50], Iter [195/209] Loss: 2.9864, average_loss: 2.4269
Epoch [42/50], Iter [200/209] Loss: 1.7645, average_loss: 2.4245
Epoch [42/50], Iter [205/209] Loss: 2.5707, average_loss: 2.4242

Starting epoch 43 / 50

Learning Rate for this epoch: 1e-05

Epoch [43/50], Iter [5/209] Loss: 1.8100, average_loss: 2.0742
Epoch [43/50], Iter [10/209] Loss: 2.3095, average_loss: 2.2729
Epoch [43/50], Iter [15/209] Loss: 2.6572, average_loss: 2.2561
Epoch [43/50], Iter [20/209] Loss: 2.3090, average_loss: 2.3479
Epoch [43/50], Iter [25/209] Loss: 2.6285, average_loss: 2.4134
Epoch [43/50], Iter [30/209] Loss: 2.2880, average_loss: 2.4206
Epoch [43/50], Iter [35/209] Loss: 2.2786, average_loss: 2.4034
Epoch [43/50], Iter [40/209] Loss: 2.1495, average_loss: 2.3899
Epoch [43/50], Iter [45/209] Loss: 2.7249, average_loss: 2.4066

Epoch [43/50], Iter [50/209] Loss: 2.5956, average_loss: 2.3987
Epoch [43/50], Iter [55/209] Loss: 2.3982, average_loss: 2.3776
Epoch [43/50], Iter [60/209] Loss: 2.2810, average_loss: 2.3869
Epoch [43/50], Iter [65/209] Loss: 1.7823, average_loss: 2.3640
Epoch [43/50], Iter [70/209] Loss: 2.0754, average_loss: 2.3507
Epoch [43/50], Iter [75/209] Loss: 4.1842, average_loss: 2.3697
Epoch [43/50], Iter [80/209] Loss: 2.3739, average_loss: 2.3808
Epoch [43/50], Iter [85/209] Loss: 3.0935, average_loss: 2.3842
Epoch [43/50], Iter [90/209] Loss: 2.6328, average_loss: 2.3759
Epoch [43/50], Iter [95/209] Loss: 2.3107, average_loss: 2.3729
Epoch [43/50], Iter [100/209] Loss: 2.6422, average_loss: 2.3999
Epoch [43/50], Iter [105/209] Loss: 2.2229, average_loss: 2.4059
Epoch [43/50], Iter [110/209] Loss: 2.0404, average_loss: 2.4013
Epoch [43/50], Iter [115/209] Loss: 2.1525, average_loss: 2.3976
Epoch [43/50], Iter [120/209] Loss: 2.5743, average_loss: 2.3884
Epoch [43/50], Iter [125/209] Loss: 2.0039, average_loss: 2.3706
Epoch [43/50], Iter [130/209] Loss: 2.7423, average_loss: 2.3778
Epoch [43/50], Iter [135/209] Loss: 3.1221, average_loss: 2.3833
Epoch [43/50], Iter [140/209] Loss: 3.3612, average_loss: 2.3919
Epoch [43/50], Iter [145/209] Loss: 1.9314, average_loss: 2.3913
Epoch [43/50], Iter [150/209] Loss: 2.2833, average_loss: 2.3871
Epoch [43/50], Iter [155/209] Loss: 2.3160, average_loss: 2.3876
Epoch [43/50], Iter [160/209] Loss: 3.2780, average_loss: 2.3798
Epoch [43/50], Iter [165/209] Loss: 2.1079, average_loss: 2.3850
Epoch [43/50], Iter [170/209] Loss: 2.6911, average_loss: 2.3959
Epoch [43/50], Iter [175/209] Loss: 2.2867, average_loss: 2.3975
Epoch [43/50], Iter [180/209] Loss: 2.1732, average_loss: 2.3999
Epoch [43/50], Iter [185/209] Loss: 2.4774, average_loss: 2.3979
Epoch [43/50], Iter [190/209] Loss: 2.9783, average_loss: 2.4062
Epoch [43/50], Iter [195/209] Loss: 2.1744, average_loss: 2.4018
Epoch [43/50], Iter [200/209] Loss: 2.7572, average_loss: 2.4012
Epoch [43/50], Iter [205/209] Loss: 2.0822, average_loss: 2.4023

Starting epoch 44 / 50

Learning Rate for this epoch: 1e-05

Epoch [44/50], Iter [5/209] Loss: 2.1533, average_loss: 2.1655
Epoch [44/50], Iter [10/209] Loss: 2.1522, average_loss: 2.3193
Epoch [44/50], Iter [15/209] Loss: 2.0143, average_loss: 2.2672
Epoch [44/50], Iter [20/209] Loss: 1.9801, average_loss: 2.1886
Epoch [44/50], Iter [25/209] Loss: 2.9370, average_loss: 2.2657
Epoch [44/50], Iter [30/209] Loss: 3.5457, average_loss: 2.3425
Epoch [44/50], Iter [35/209] Loss: 2.1550, average_loss: 2.3289
Epoch [44/50], Iter [40/209] Loss: 2.1087, average_loss: 2.2759
Epoch [44/50], Iter [45/209] Loss: 2.3075, average_loss: 2.2695
Epoch [44/50], Iter [50/209] Loss: 1.9165, average_loss: 2.2542
Epoch [44/50], Iter [55/209] Loss: 2.0380, average_loss: 2.2387
Epoch [44/50], Iter [60/209] Loss: 2.7183, average_loss: 2.2776
Epoch [44/50], Iter [65/209] Loss: 1.7736, average_loss: 2.2679
Epoch [44/50], Iter [70/209] Loss: 2.4853, average_loss: 2.2633
Epoch [44/50], Iter [75/209] Loss: 2.5248, average_loss: 2.2791
Epoch [44/50], Iter [80/209] Loss: 2.3526, average_loss: 2.2945
Epoch [44/50], Iter [85/209] Loss: 2.2303, average_loss: 2.3076
Epoch [44/50], Iter [90/209] Loss: 1.6099, average_loss: 2.3120

Epoch [44/50], Iter [95/209] Loss: 2.8935, average_loss: 2.3398
Epoch [44/50], Iter [100/209] Loss: 2.7453, average_loss: 2.3427
Epoch [44/50], Iter [105/209] Loss: 2.8806, average_loss: 2.3529
Epoch [44/50], Iter [110/209] Loss: 2.2217, average_loss: 2.3407
Epoch [44/50], Iter [115/209] Loss: 2.5066, average_loss: 2.3511
Epoch [44/50], Iter [120/209] Loss: 1.7129, average_loss: 2.3501
Epoch [44/50], Iter [125/209] Loss: 2.9276, average_loss: 2.3551
Epoch [44/50], Iter [130/209] Loss: 2.1274, average_loss: 2.3580
Epoch [44/50], Iter [135/209] Loss: 2.2092, average_loss: 2.3595
Epoch [44/50], Iter [140/209] Loss: 2.3680, average_loss: 2.3801
Epoch [44/50], Iter [145/209] Loss: 3.1641, average_loss: 2.3811
Epoch [44/50], Iter [150/209] Loss: 3.2331, average_loss: 2.3779
Epoch [44/50], Iter [155/209] Loss: 2.5033, average_loss: 2.3717
Epoch [44/50], Iter [160/209] Loss: 2.0902, average_loss: 2.3684
Epoch [44/50], Iter [165/209] Loss: 1.8057, average_loss: 2.3684
Epoch [44/50], Iter [170/209] Loss: 2.4051, average_loss: 2.3796
Epoch [44/50], Iter [175/209] Loss: 3.5379, average_loss: 2.3939
Epoch [44/50], Iter [180/209] Loss: 2.5067, average_loss: 2.3986
Epoch [44/50], Iter [185/209] Loss: 2.5306, average_loss: 2.3957
Epoch [44/50], Iter [190/209] Loss: 2.8740, average_loss: 2.4019
Epoch [44/50], Iter [195/209] Loss: 3.2067, average_loss: 2.4016
Epoch [44/50], Iter [200/209] Loss: 3.0216, average_loss: 2.4081
Epoch [44/50], Iter [205/209] Loss: 2.3282, average_loss: 2.4113

Starting epoch 45 / 50

Learning Rate for this epoch: 1e-05

Epoch [45/50], Iter [5/209] Loss: 2.7224, average_loss: 2.3463
Epoch [45/50], Iter [10/209] Loss: 2.0704, average_loss: 2.3231
Epoch [45/50], Iter [15/209] Loss: 1.6812, average_loss: 2.3017
Epoch [45/50], Iter [20/209] Loss: 2.4338, average_loss: 2.3298
Epoch [45/50], Iter [25/209] Loss: 2.8614, average_loss: 2.3437
Epoch [45/50], Iter [30/209] Loss: 2.9057, average_loss: 2.3834
Epoch [45/50], Iter [35/209] Loss: 2.4439, average_loss: 2.3582
Epoch [45/50], Iter [40/209] Loss: 2.3066, average_loss: 2.3745
Epoch [45/50], Iter [45/209] Loss: 2.6484, average_loss: 2.3502
Epoch [45/50], Iter [50/209] Loss: 2.9339, average_loss: 2.3681
Epoch [45/50], Iter [55/209] Loss: 2.4792, average_loss: 2.3684
Epoch [45/50], Iter [60/209] Loss: 2.5773, average_loss: 2.3755
Epoch [45/50], Iter [65/209] Loss: 1.9052, average_loss: 2.3755
Epoch [45/50], Iter [70/209] Loss: 1.7766, average_loss: 2.3651
Epoch [45/50], Iter [75/209] Loss: 2.5146, average_loss: 2.3748
Epoch [45/50], Iter [80/209] Loss: 2.9407, average_loss: 2.3828
Epoch [45/50], Iter [85/209] Loss: 2.5020, average_loss: 2.3839
Epoch [45/50], Iter [90/209] Loss: 2.5239, average_loss: 2.3814
Epoch [45/50], Iter [95/209] Loss: 1.9680, average_loss: 2.3823
Epoch [45/50], Iter [100/209] Loss: 2.1098, average_loss: 2.3616
Epoch [45/50], Iter [105/209] Loss: 2.6458, average_loss: 2.3645
Epoch [45/50], Iter [110/209] Loss: 2.4290, average_loss: 2.3723
Epoch [45/50], Iter [115/209] Loss: 1.8020, average_loss: 2.3712
Epoch [45/50], Iter [120/209] Loss: 2.2090, average_loss: 2.3867
Epoch [45/50], Iter [125/209] Loss: 2.9211, average_loss: 2.3814
Epoch [45/50], Iter [130/209] Loss: 2.1660, average_loss: 2.3753
Epoch [45/50], Iter [135/209] Loss: 2.6470, average_loss: 2.3909

Epoch [45/50], Iter [140/209] Loss: 2.3743, average_loss: 2.3887
Epoch [45/50], Iter [145/209] Loss: 2.1908, average_loss: 2.3918
Epoch [45/50], Iter [150/209] Loss: 2.3937, average_loss: 2.3994
Epoch [45/50], Iter [155/209] Loss: 2.6300, average_loss: 2.4127
Epoch [45/50], Iter [160/209] Loss: 3.2625, average_loss: 2.4176
Epoch [45/50], Iter [165/209] Loss: 2.3056, average_loss: 2.4278
Epoch [45/50], Iter [170/209] Loss: 1.9959, average_loss: 2.4357
Epoch [45/50], Iter [175/209] Loss: 2.3760, average_loss: 2.4398
Epoch [45/50], Iter [180/209] Loss: 1.5615, average_loss: 2.4389
Epoch [45/50], Iter [185/209] Loss: 2.2624, average_loss: 2.4315
Epoch [45/50], Iter [190/209] Loss: 1.5055, average_loss: 2.4395
Epoch [45/50], Iter [195/209] Loss: 2.2501, average_loss: 2.4390
Epoch [45/50], Iter [200/209] Loss: 2.4244, average_loss: 2.4434
Epoch [45/50], Iter [205/209] Loss: 2.3997, average_loss: 2.4464
Updating best test loss: 3.47855

Starting epoch 46 / 50

Learning Rate for this epoch: 1e-05

Epoch [46/50], Iter [5/209] Loss: 2.1555, average_loss: 2.3147
Epoch [46/50], Iter [10/209] Loss: 2.5044, average_loss: 2.3699
Epoch [46/50], Iter [15/209] Loss: 3.6795, average_loss: 2.3919
Epoch [46/50], Iter [20/209] Loss: 2.3254, average_loss: 2.3296
Epoch [46/50], Iter [25/209] Loss: 1.8654, average_loss: 2.2728
Epoch [46/50], Iter [30/209] Loss: 2.7397, average_loss: 2.2868
Epoch [46/50], Iter [35/209] Loss: 2.3936, average_loss: 2.2469
Epoch [46/50], Iter [40/209] Loss: 2.2864, average_loss: 2.2476
Epoch [46/50], Iter [45/209] Loss: 2.7540, average_loss: 2.2941
Epoch [46/50], Iter [50/209] Loss: 2.9264, average_loss: 2.3013
Epoch [46/50], Iter [55/209] Loss: 2.4319, average_loss: 2.3066
Epoch [46/50], Iter [60/209] Loss: 2.5036, average_loss: 2.2831
Epoch [46/50], Iter [65/209] Loss: 1.9502, average_loss: 2.2800
Epoch [46/50], Iter [70/209] Loss: 2.8890, average_loss: 2.3155
Epoch [46/50], Iter [75/209] Loss: 2.7729, average_loss: 2.3399
Epoch [46/50], Iter [80/209] Loss: 2.1489, average_loss: 2.3443
Epoch [46/50], Iter [85/209] Loss: 2.2490, average_loss: 2.3670
Epoch [46/50], Iter [90/209] Loss: 2.2869, average_loss: 2.3863
Epoch [46/50], Iter [95/209] Loss: 2.2213, average_loss: 2.3880
Epoch [46/50], Iter [100/209] Loss: 2.6399, average_loss: 2.3791
Epoch [46/50], Iter [105/209] Loss: 2.2738, average_loss: 2.3645
Epoch [46/50], Iter [110/209] Loss: 2.4338, average_loss: 2.3610
Epoch [46/50], Iter [115/209] Loss: 1.7666, average_loss: 2.3612
Epoch [46/50], Iter [120/209] Loss: 2.2882, average_loss: 2.3511
Epoch [46/50], Iter [125/209] Loss: 2.2476, average_loss: 2.3549
Epoch [46/50], Iter [130/209] Loss: 2.6796, average_loss: 2.3740
Epoch [46/50], Iter [135/209] Loss: 2.4201, average_loss: 2.3864
Epoch [46/50], Iter [140/209] Loss: 2.7770, average_loss: 2.4008
Epoch [46/50], Iter [145/209] Loss: 1.8518, average_loss: 2.4030
Epoch [46/50], Iter [150/209] Loss: 2.6633, average_loss: 2.3986
Epoch [46/50], Iter [155/209] Loss: 3.3589, average_loss: 2.4088
Epoch [46/50], Iter [160/209] Loss: 2.5568, average_loss: 2.4042
Epoch [46/50], Iter [165/209] Loss: 2.3780, average_loss: 2.4049
Epoch [46/50], Iter [170/209] Loss: 2.3844, average_loss: 2.4168
Epoch [46/50], Iter [175/209] Loss: 2.7421, average_loss: 2.4285

Epoch [46/50], Iter [180/209] Loss: 2.1751, average_loss: 2.4302
Epoch [46/50], Iter [185/209] Loss: 1.8327, average_loss: 2.4230
Epoch [46/50], Iter [190/209] Loss: 3.2150, average_loss: 2.4289
Epoch [46/50], Iter [195/209] Loss: 2.3833, average_loss: 2.4264
Epoch [46/50], Iter [200/209] Loss: 3.2902, average_loss: 2.4296
Epoch [46/50], Iter [205/209] Loss: 2.8825, average_loss: 2.4259

Starting epoch 47 / 50

Learning Rate for this epoch: 1e-05

Epoch [47/50], Iter [5/209] Loss: 1.4762, average_loss: 2.1774
Epoch [47/50], Iter [10/209] Loss: 2.8033, average_loss: 2.1864
Epoch [47/50], Iter [15/209] Loss: 2.3757, average_loss: 2.2258
Epoch [47/50], Iter [20/209] Loss: 2.3992, average_loss: 2.2782
Epoch [47/50], Iter [25/209] Loss: 2.2744, average_loss: 2.3268
Epoch [47/50], Iter [30/209] Loss: 2.7860, average_loss: 2.4018
Epoch [47/50], Iter [35/209] Loss: 2.7104, average_loss: 2.4167
Epoch [47/50], Iter [40/209] Loss: 2.5344, average_loss: 2.4007
Epoch [47/50], Iter [45/209] Loss: 1.9165, average_loss: 2.3708
Epoch [47/50], Iter [50/209] Loss: 2.5459, average_loss: 2.4130
Epoch [47/50], Iter [55/209] Loss: 1.6741, average_loss: 2.3768
Epoch [47/50], Iter [60/209] Loss: 2.2649, average_loss: 2.3831
Epoch [47/50], Iter [65/209] Loss: 2.0593, average_loss: 2.3890
Epoch [47/50], Iter [70/209] Loss: 2.8315, average_loss: 2.3833
Epoch [47/50], Iter [75/209] Loss: 2.9046, average_loss: 2.3887
Epoch [47/50], Iter [80/209] Loss: 2.7333, average_loss: 2.4008
Epoch [47/50], Iter [85/209] Loss: 2.6334, average_loss: 2.4072
Epoch [47/50], Iter [90/209] Loss: 1.9705, average_loss: 2.4086
Epoch [47/50], Iter [95/209] Loss: 1.9192, average_loss: 2.4028
Epoch [47/50], Iter [100/209] Loss: 1.9625, average_loss: 2.4000
Epoch [47/50], Iter [105/209] Loss: 2.7537, average_loss: 2.4156
Epoch [47/50], Iter [110/209] Loss: 2.4694, average_loss: 2.4107
Epoch [47/50], Iter [115/209] Loss: 2.5697, average_loss: 2.4056
Epoch [47/50], Iter [120/209] Loss: 3.0129, average_loss: 2.4012
Epoch [47/50], Iter [125/209] Loss: 2.1950, average_loss: 2.4001
Epoch [47/50], Iter [130/209] Loss: 1.9284, average_loss: 2.3941
Epoch [47/50], Iter [135/209] Loss: 2.1324, average_loss: 2.3919
Epoch [47/50], Iter [140/209] Loss: 2.7489, average_loss: 2.4110
Epoch [47/50], Iter [145/209] Loss: 1.7122, average_loss: 2.3975
Epoch [47/50], Iter [150/209] Loss: 1.7055, average_loss: 2.3996
Epoch [47/50], Iter [155/209] Loss: 2.1936, average_loss: 2.4080
Epoch [47/50], Iter [160/209] Loss: 2.4032, average_loss: 2.4074
Epoch [47/50], Iter [165/209] Loss: 2.4517, average_loss: 2.4063
Epoch [47/50], Iter [170/209] Loss: 2.2100, average_loss: 2.4054
Epoch [47/50], Iter [175/209] Loss: 1.6524, average_loss: 2.4079
Epoch [47/50], Iter [180/209] Loss: 3.1113, average_loss: 2.4057
Epoch [47/50], Iter [185/209] Loss: 2.5839, average_loss: 2.4017
Epoch [47/50], Iter [190/209] Loss: 2.5967, average_loss: 2.4031
Epoch [47/50], Iter [195/209] Loss: 2.5942, average_loss: 2.4006
Epoch [47/50], Iter [200/209] Loss: 1.2808, average_loss: 2.3913
Epoch [47/50], Iter [205/209] Loss: 2.0159, average_loss: 2.3833

Starting epoch 48 / 50

Learning Rate for this epoch: 1e-05

Epoch [48/50], Iter [5/209] Loss: 1.9046, average_loss: 2.3067
Epoch [48/50], Iter [10/209] Loss: 1.9732, average_loss: 2.4276
Epoch [48/50], Iter [15/209] Loss: 2.0254, average_loss: 2.4180
Epoch [48/50], Iter [20/209] Loss: 2.2096, average_loss: 2.3928
Epoch [48/50], Iter [25/209] Loss: 2.0846, average_loss: 2.4074
Epoch [48/50], Iter [30/209] Loss: 2.8181, average_loss: 2.4524
Epoch [48/50], Iter [35/209] Loss: 1.6591, average_loss: 2.3858
Epoch [48/50], Iter [40/209] Loss: 2.3351, average_loss: 2.3612
Epoch [48/50], Iter [45/209] Loss: 2.4497, average_loss: 2.3257
Epoch [48/50], Iter [50/209] Loss: 1.9232, average_loss: 2.3209
Epoch [48/50], Iter [55/209] Loss: 1.6417, average_loss: 2.3343
Epoch [48/50], Iter [60/209] Loss: 2.2171, average_loss: 2.3367
Epoch [48/50], Iter [65/209] Loss: 1.8746, average_loss: 2.3148
Epoch [48/50], Iter [70/209] Loss: 2.4747, average_loss: 2.3432
Epoch [48/50], Iter [75/209] Loss: 2.6956, average_loss: 2.3411
Epoch [48/50], Iter [80/209] Loss: 2.7988, average_loss: 2.3508
Epoch [48/50], Iter [85/209] Loss: 2.7252, average_loss: 2.3679
Epoch [48/50], Iter [90/209] Loss: 1.6749, average_loss: 2.3662
Epoch [48/50], Iter [95/209] Loss: 2.5101, average_loss: 2.3698
Epoch [48/50], Iter [100/209] Loss: 2.6242, average_loss: 2.3656
Epoch [48/50], Iter [105/209] Loss: 3.7099, average_loss: 2.3692
Epoch [48/50], Iter [110/209] Loss: 3.1149, average_loss: 2.3838
Epoch [48/50], Iter [115/209] Loss: 2.5265, average_loss: 2.3808
Epoch [48/50], Iter [120/209] Loss: 2.5037, average_loss: 2.3745
Epoch [48/50], Iter [125/209] Loss: 2.3467, average_loss: 2.3809
Epoch [48/50], Iter [130/209] Loss: 2.5852, average_loss: 2.3824
Epoch [48/50], Iter [135/209] Loss: 2.6616, average_loss: 2.3817
Epoch [48/50], Iter [140/209] Loss: 1.5169, average_loss: 2.3728
Epoch [48/50], Iter [145/209] Loss: 1.6686, average_loss: 2.3757
Epoch [48/50], Iter [150/209] Loss: 2.2950, average_loss: 2.3697
Epoch [48/50], Iter [155/209] Loss: 2.8508, average_loss: 2.3746
Epoch [48/50], Iter [160/209] Loss: 3.1487, average_loss: 2.3794
Epoch [48/50], Iter [165/209] Loss: 1.7099, average_loss: 2.3744
Epoch [48/50], Iter [170/209] Loss: 3.1520, average_loss: 2.3781
Epoch [48/50], Iter [175/209] Loss: 1.6368, average_loss: 2.3727
Epoch [48/50], Iter [180/209] Loss: 2.0018, average_loss: 2.3825
Epoch [48/50], Iter [185/209] Loss: 2.4592, average_loss: 2.3839
Epoch [48/50], Iter [190/209] Loss: 3.4363, average_loss: 2.3904
Epoch [48/50], Iter [195/209] Loss: 2.2171, average_loss: 2.3859
Epoch [48/50], Iter [200/209] Loss: 2.5472, average_loss: 2.3844
Epoch [48/50], Iter [205/209] Loss: 2.1141, average_loss: 2.3833

Starting epoch 49 / 50

Learning Rate for this epoch: 1e-05

Epoch [49/50], Iter [5/209] Loss: 2.5582, average_loss: 2.5736
Epoch [49/50], Iter [10/209] Loss: 2.9344, average_loss: 2.4871
Epoch [49/50], Iter [15/209] Loss: 1.8744, average_loss: 2.4229
Epoch [49/50], Iter [20/209] Loss: 2.1894, average_loss: 2.3663
Epoch [49/50], Iter [25/209] Loss: 2.0474, average_loss: 2.3267
Epoch [49/50], Iter [30/209] Loss: 2.4673, average_loss: 2.3785
Epoch [49/50], Iter [35/209] Loss: 2.0063, average_loss: 2.3747
Epoch [49/50], Iter [40/209] Loss: 2.1708, average_loss: 2.3784

Epoch [49/50], Iter [45/209] Loss: 2.2400, average_loss: 2.3594
Epoch [49/50], Iter [50/209] Loss: 2.4204, average_loss: 2.3555
Epoch [49/50], Iter [55/209] Loss: 2.4928, average_loss: 2.3944
Epoch [49/50], Iter [60/209] Loss: 1.9917, average_loss: 2.3672
Epoch [49/50], Iter [65/209] Loss: 2.5598, average_loss: 2.3490
Epoch [49/50], Iter [70/209] Loss: 2.9144, average_loss: 2.3686
Epoch [49/50], Iter [75/209] Loss: 1.5335, average_loss: 2.3726
Epoch [49/50], Iter [80/209] Loss: 2.0524, average_loss: 2.3678
Epoch [49/50], Iter [85/209] Loss: 3.3556, average_loss: 2.3583
Epoch [49/50], Iter [90/209] Loss: 2.8062, average_loss: 2.3782
Epoch [49/50], Iter [95/209] Loss: 2.3359, average_loss: 2.3696
Epoch [49/50], Iter [100/209] Loss: 1.9301, average_loss: 2.3687
Epoch [49/50], Iter [105/209] Loss: 2.3113, average_loss: 2.3612
Epoch [49/50], Iter [110/209] Loss: 2.5774, average_loss: 2.3720
Epoch [49/50], Iter [115/209] Loss: 2.1154, average_loss: 2.3674
Epoch [49/50], Iter [120/209] Loss: 2.5656, average_loss: 2.3695
Epoch [49/50], Iter [125/209] Loss: 1.9770, average_loss: 2.3901
Epoch [49/50], Iter [130/209] Loss: 2.3380, average_loss: 2.3910
Epoch [49/50], Iter [135/209] Loss: 1.8495, average_loss: 2.3859
Epoch [49/50], Iter [140/209] Loss: 1.6343, average_loss: 2.3829
Epoch [49/50], Iter [145/209] Loss: 1.9875, average_loss: 2.3875
Epoch [49/50], Iter [150/209] Loss: 2.3415, average_loss: 2.3939
Epoch [49/50], Iter [155/209] Loss: 1.8848, average_loss: 2.3954
Epoch [49/50], Iter [160/209] Loss: 3.0348, average_loss: 2.4000
Epoch [49/50], Iter [165/209] Loss: 2.7050, average_loss: 2.4013
Epoch [49/50], Iter [170/209] Loss: 2.1473, average_loss: 2.4059
Epoch [49/50], Iter [175/209] Loss: 1.7153, average_loss: 2.4029
Epoch [49/50], Iter [180/209] Loss: 2.4747, average_loss: 2.4039
Epoch [49/50], Iter [185/209] Loss: 2.7166, average_loss: 2.3957
Epoch [49/50], Iter [190/209] Loss: 2.9085, average_loss: 2.3976
Epoch [49/50], Iter [195/209] Loss: 2.9116, average_loss: 2.4050
Epoch [49/50], Iter [200/209] Loss: 2.1842, average_loss: 2.4109
Epoch [49/50], Iter [205/209] Loss: 1.8848, average_loss: 2.4083

Starting epoch 50 / 50

Learning Rate for this epoch: 1e-05

Epoch [50/50], Iter [5/209] Loss: 2.7043, average_loss: 2.4802
Epoch [50/50], Iter [10/209] Loss: 2.4313, average_loss: 2.4048
Epoch [50/50], Iter [15/209] Loss: 2.9494, average_loss: 2.3677
Epoch [50/50], Iter [20/209] Loss: 2.1309, average_loss: 2.2999
Epoch [50/50], Iter [25/209] Loss: 2.7784, average_loss: 2.4108
Epoch [50/50], Iter [30/209] Loss: 3.2521, average_loss: 2.4275
Epoch [50/50], Iter [35/209] Loss: 2.7526, average_loss: 2.4205
Epoch [50/50], Iter [40/209] Loss: 1.9307, average_loss: 2.3787
Epoch [50/50], Iter [45/209] Loss: 2.3794, average_loss: 2.3896
Epoch [50/50], Iter [50/209] Loss: 3.1557, average_loss: 2.3941
Epoch [50/50], Iter [55/209] Loss: 1.9465, average_loss: 2.3980
Epoch [50/50], Iter [60/209] Loss: 1.7929, average_loss: 2.3860
Epoch [50/50], Iter [65/209] Loss: 2.8074, average_loss: 2.3961
Epoch [50/50], Iter [70/209] Loss: 2.6228, average_loss: 2.4079
Epoch [50/50], Iter [75/209] Loss: 2.8250, average_loss: 2.3994
Epoch [50/50], Iter [80/209] Loss: 2.4195, average_loss: 2.3930
Epoch [50/50], Iter [85/209] Loss: 2.6309, average_loss: 2.4010

Epoch [50/50], Iter [90/209] Loss: 1.9126, average_loss: 2.3887
Epoch [50/50], Iter [95/209] Loss: 2.7497, average_loss: 2.3870
Epoch [50/50], Iter [100/209] Loss: 3.1336, average_loss: 2.3994
Epoch [50/50], Iter [105/209] Loss: 2.3771, average_loss: 2.3846
Epoch [50/50], Iter [110/209] Loss: 2.2356, average_loss: 2.3817
Epoch [50/50], Iter [115/209] Loss: 2.7191, average_loss: 2.3814
Epoch [50/50], Iter [120/209] Loss: 2.4544, average_loss: 2.3975
Epoch [50/50], Iter [125/209] Loss: 1.9755, average_loss: 2.3973
Epoch [50/50], Iter [130/209] Loss: 2.8287, average_loss: 2.4006
Epoch [50/50], Iter [135/209] Loss: 2.3450, average_loss: 2.4058
Epoch [50/50], Iter [140/209] Loss: 4.1643, average_loss: 2.4114
Epoch [50/50], Iter [145/209] Loss: 1.8806, average_loss: 2.4024
Epoch [50/50], Iter [150/209] Loss: 3.0991, average_loss: 2.4198
Epoch [50/50], Iter [155/209] Loss: 2.6819, average_loss: 2.4179
Epoch [50/50], Iter [160/209] Loss: 2.4447, average_loss: 2.4140
Epoch [50/50], Iter [165/209] Loss: 2.2021, average_loss: 2.4202
Epoch [50/50], Iter [170/209] Loss: 3.5617, average_loss: 2.4182
Epoch [50/50], Iter [175/209] Loss: 2.1411, average_loss: 2.4147
Epoch [50/50], Iter [180/209] Loss: 2.9357, average_loss: 2.4203
Epoch [50/50], Iter [185/209] Loss: 1.9828, average_loss: 2.4203
Epoch [50/50], Iter [190/209] Loss: 2.4957, average_loss: 2.4265
Epoch [50/50], Iter [195/209] Loss: 3.0747, average_loss: 2.4334
Epoch [50/50], Iter [200/209] Loss: 1.7174, average_loss: 2.4325
Epoch [50/50], Iter [205/209] Loss: 3.0214, average_loss: 2.4353

---Evaluate model on test samples---

---class aeroplane ap 0.5630749983513839---
---class bicycle ap 0.6799799142283313---
---class bird ap 0.5541332648547633---
---class boat ap 0.3569283653117547---
---class bottle ap 0.3070483305863663---
---class bus ap 0.6793249944180304---
---class car ap 0.7231935630522555---
---class cat ap 0.7584718170262736---
---class chair ap 0.3632634345867318---
---class cow ap 0.5826046381409622---
---class diningtable ap 0.41294517094080785---
---class dog ap 0.6613101916215132---
---class horse ap 0.6946512670404343---
---class motorbike ap 0.6596461476034197---
---class person ap 0.6239706982984516---
---class pottedplant ap 0.2517757422116694---
---class sheep ap 0.50424957299758---
---class sofa ap 0.5133728725446116---
---class train ap 0.7194507667291103---
---class tvmonitor ap 0.5734353361682977---
---map 0.5591415543356375---