In [ ]:
```
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_loss2 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
```

In [2]:
```
# !bash download_data.sh
```

Assignment3 Part2: Yolo Detection

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

What to do

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

What to submit

See the submission template for what to submit.

Initialization

In [3]:
```
device = torch.device("cuda:0" if torch.cuda.is_available() else "cpu")
```

In [4]:
```
# 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 w
```

To implement Yolo we will rely on a pretrained classifier as the backbone for our detection network.

PyTorch offers a variety of models which are pretrained on ImageNet in the `torchvision.models` (https://pytorch.org/docs/stable/torchvision/models.html) package. In particular, we will use the ResNet50 architecture as a base for our detector. This is different from the base architecture in the Yolo paper and also results in a different output grid size (14x14 instead of 7x7).

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

In [5]:
```
load_network_path = "detector.pth"
# load_network_path = None
pretrained = True

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

```
Loading saved network from detector.pth
```

In [6]:
```
learning_rate = 0.001
num_epochs = 50
batch_size = 24

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

In [7]:
```
criterion = YoloLoss(S, B, lambda_coord, lambda_noobj)
optimizer = torch.optim.SGD(net.parameters(), lr=learning_rate, momentum=0.9, wei
```

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

In [8]:
```
file_root_train = 'VOCdevkit_2007/VOC2007/JPEGImages/'
annotation_file_train = 'voc2007.txt'

train_dataset = VocDetectorDataset(root_img_dir=file_root_train,dataset_file=anno
train_loader = DataLoader(train_dataset,batch_size=batch_size,shuffle=True,num_wo
print('Loaded %d train images' % len(train_dataset))
```

```
Initializing dataset
Loaded 5011 train images
```

In [9]:
```
file_root_test = 'VOCdevkit_2007/VOC2007test/JPEGImages/'
annotation_file_test = 'voc2007test.txt'

test_dataset = VocDetectorDataset(root_img_dir=file_root_test,dataset_file=annota
test_loader = DataLoader(test_dataset,batch_size=batch_size,shuffle=False,num_wor
print('Loaded %d test images' % len(test_dataset))
```

```
Initializing dataset
Loaded 4950 test images
```

In [10]:
```
import warnings
warnings.filterwarnings("ignore")
```

Train detector

In [11]:

```
best_test_loss = np.inf

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

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

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

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

    total_loss = 0.

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

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

        optimizer.zero_grad()
        loss.backward()
        optimizer.step()
        if (i+1) % 5 == 0:
            print('Epoch [%d/%d], Iter [%d/%d] Loss: %.4f, average_loss: %.4f'
                  % (epoch+1, num_epochs, i+1, len(train_loader), loss.item(), t

    # 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')
#     if(epoch+1%10==0):
#         test_aps = evaluate(net, test_dataset_file=annotation_file_test)

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

```
Epoch [29/30], Iter [200/209] Loss: 1.2037, average_loss: 1.7339
Epoch [29/30], Iter [205/209] Loss: 1.6568, average_loss: 1.7326
```

```
Starting epoch 30 / 30
Learning Rate for this epoch: 0.001
Epoch [30/30], Iter [5/209] Loss: 2.0015, average_loss: 1.9317
Epoch [30/30], Iter [10/209] Loss: 1.6260, average_loss: 1.8807
Epoch [30/30], Iter [15/209] Loss: 2.4334, average_loss: 1.8883
Epoch [30/30], Iter [20/209] Loss: 1.2987, average_loss: 1.7650
Epoch [30/30], Iter [25/209] Loss: 1.7944, average_loss: 1.7779
Epoch [30/30], Iter [30/209] Loss: 1.6117, average_loss: 1.7560
Epoch [30/30], Iter [35/209] Loss: 1.3166, average_loss: 1.7422
Epoch [30/30], Iter [40/209] Loss: 1.2474, average_loss: 1.7144
Epoch [30/30], Iter [45/209] Loss: 1.4536, average_loss: 1.7073
Epoch [30/30], Iter [50/209] Loss: 1.3129, average_loss: 1.6812
Epoch [30/30], Iter [55/209] Loss: 1.1086, average_loss: 1.6565
Epoch [30/30], Iter [60/209] Loss: 1.6351, average_loss: 1.6553
Epoch [30/30], Iter [65/209] Loss: 1.7209, average_loss: 1.6784
Epoch [30/30], Iter [70/209] Loss: 1.8686, average_loss: 1.6749
```

View example predictions

In [12]:
```
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_si
                  color, -1)
    cv2.putText(image, label, (p1[0], p1[1] + baseline), cv2.FONT_HERSHEY_SIMPLEX

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

```
predicting...
```

Out[12]: <matplotlib.image.AxesImage at 0x7f49c56a9dd0>

Evaluate on Test

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

In [13]:
```
test_aps = evaluate(net, test_dataset_file=annotation_file_test)
```

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

100%|██████████| 4950/4950 [05:46<00:00, 14.27it/s]

---class aeroplane ap 0.5584034970809064---
---class bicycle ap 0.5499822433704729---
---class bird ap 0.5193139085156483---
---class boat ap 0.3208293760195095---
---class bottle ap 0.24913155638809353---
---class bus ap 0.6315389693721976---
---class car ap 0.6955809599311814---
---class cat ap 0.7139134069160338---
---class chair ap 0.29152022647386855---
---class cow ap 0.4763657704933756---
---class diningtable ap 0.38079999540309734---
---class dog ap 0.6712307904418868---
---class horse ap 0.6800598160418064---
---class motorbike ap 0.4998163730735388---
---class person ap 0.5477949913313129---
---class pottedplant ap 0.2151721803230729---
---class sheep ap 0.4671697883264807---
---class sofa ap 0.42021797642049596---
---class train ap 0.6184166756452507---
---class tvmonitor ap 0.49782883615182405---
---map 0.5002543668860026---
```

In [14]:
```
output_submission_csv('my_solution.csv', test_aps)
```

In [ ]:
```

```

Present | Slides | Themes | Help