Computer Vision
CSCI-UA.0480-001 2020
Assignment 6

April 4, 2020

Overview

This assignment requires you to train a Faster R-CNN model to perform object detection with the EgoHands dataset. There are mainly four steps: (i) fill in the blank in starter code to create a complete dataloader; (ii) create a Faster R-CNN model; (iii) write a complete training and validation pipeline; (iv) write a evaluate.py to evaluate your trained Faster R-CNN.

Object Detection

Here is the official torchvision object detection finetunning tutorial. Before you start working on this assignment, complete the tutorial to understand how to do object detection in pytorch. This tutorial used Mask R-CNN which is a better version of Faster R-CNN but for this assignment you only need to use Faster R-CNN. The only major difference between Mask R-CNN and Faster R-CNN is that Mask R-CNN also outputs the complete mask of the input image. You are allowed to use the engine.py mentioned in the tutorial. You should save the training logs to a pdf file.

After you finish your main.py which trains a Faster R-CNN, write a evaluate.py which loads your trained Faster R-CNN, randomly select 5 images in Egohands and save these images with predicted bounding boxes.

EgoHands

EgoHands dataset contains in total 4800 labeled images from 48 Google Glass videos. The ground-truth labels consist of polygons with hand segmentations and are provided as Matlab files. In the starter code, all polygons have been converted to masks. If you are not familiar with masks, here is an example. In the example image, all black pixels in the original mask are 0 whereas all all yellow pixels are 1. You should use these masks to compute all minimal bounding boxes. Each bounding box corresponds to one object (hand) and the label for them should all be 1 (since 0 is for background). Fig 1. is an example of minimal bounding boxes.

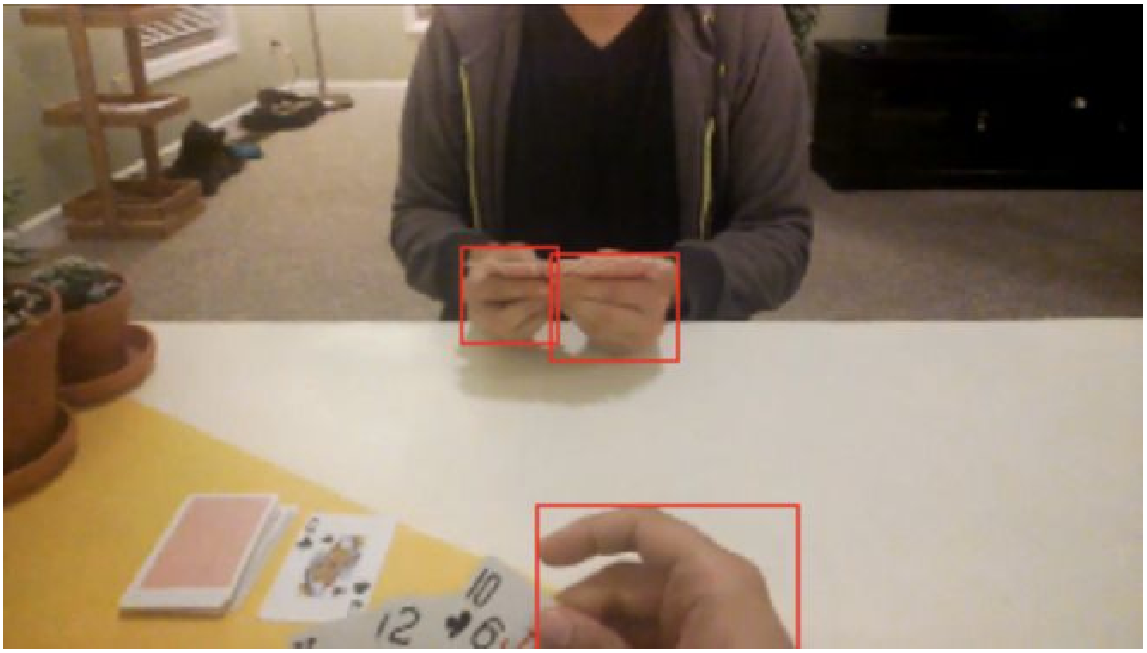

Figure 1. Bounding box example in EgoHands

Use these object detection bounding boxes along with image data to train the Faster R-CNN. Notice that some images might not contain any hand at all and you might want to omit these images. There is no train/val split in the original dataset and you should split it yourself.

Submit your scripts, report and model file to nyu classes. This assignment is due on **Saturday April 18th**.

Grading

Deliverables	Points
Dataloader	20
Faster R-CNN	10
Training and validation	20
Evaluation	20
Total	70