DATA.ML.300 Computer Vision Exercise 3

Trung Nguyen

January 2024

1

Some result of the Histogram of Gradient

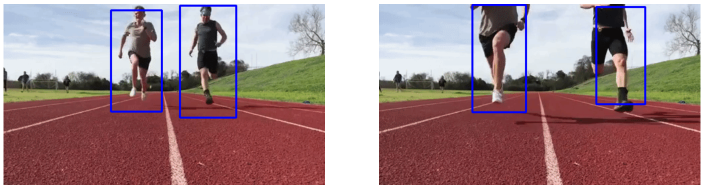

Figure 1: Person detection result using HoG

2

2.1. What is the form the targets are presented in? What is the difference between training and validation datasets in a general sense?
Target data has format (frame, xmin, xmax, ymin, ymax, class_id):
- frame: name of frame
- xmin: bounding box top x coordinate
- xmax: bounding box bottom x coordinate
- ymin: bounding box top y coordinate
- ymax: bounding box bottom y coordinate
- class_id: object id

The validation dataset has less data than training set

2.2. How many convolutional ”blocks” are there, and what kind of layers is each block build from?
- There are 7 convolutional blocks in the network before 4 predictors. Each convolutional block contains, 1 convolutional layer, 1 batch normalization layer, 1 ELU activation layer and 1 max pooling 2D layer.

2.3. What are the two attributes this loss function observes? How are these defined (short explanation without any formulas is sufficient)?

- The loss function combines the localization loss and the confidence loss.
- For localization loss, the author uses the Smooth L1 (absolute) loss between the ground truth box and the prediction. In which, the predicted box will be compared to the ground truth using 4 values, the center of the box (cx, cy) and the size of the box (width, height)
- The confidence loss is the classification loss over multiple classes, which uses softmax log loss.

3

Some webcam detection results of using SSD300

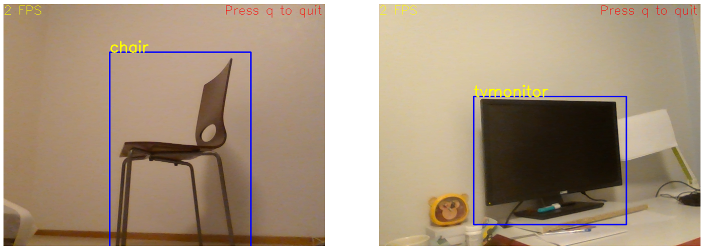

Figure 2: Object detection results of SSD300