DATA.ML.300 Computer Vision Exercise 3

Uyen Phan

01 2024

1

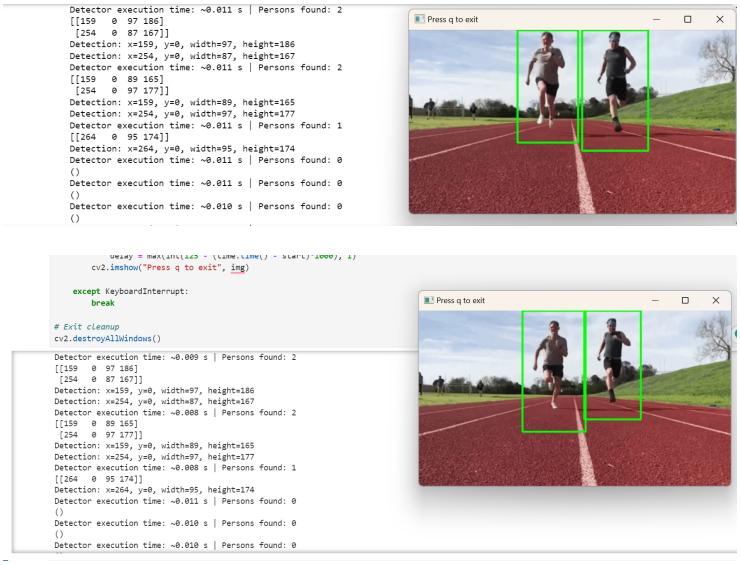


Figure 1: People detector using HOG

2

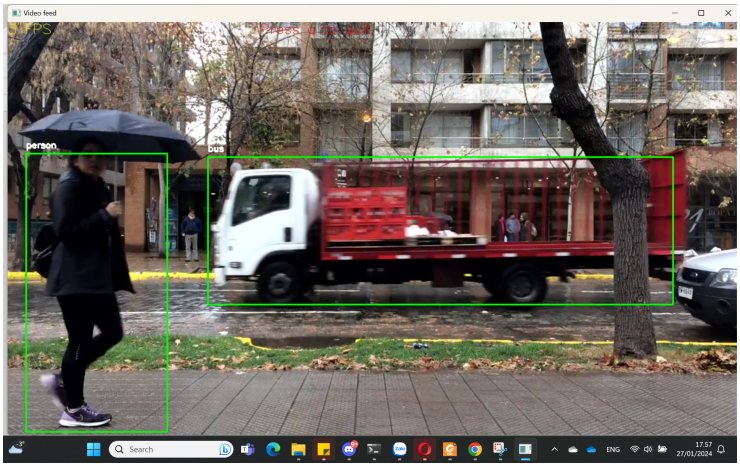
a) Targets are presented in a tabular structure where each row corresponds to an annotation for a specific image, and columns contain information about the detected objects.

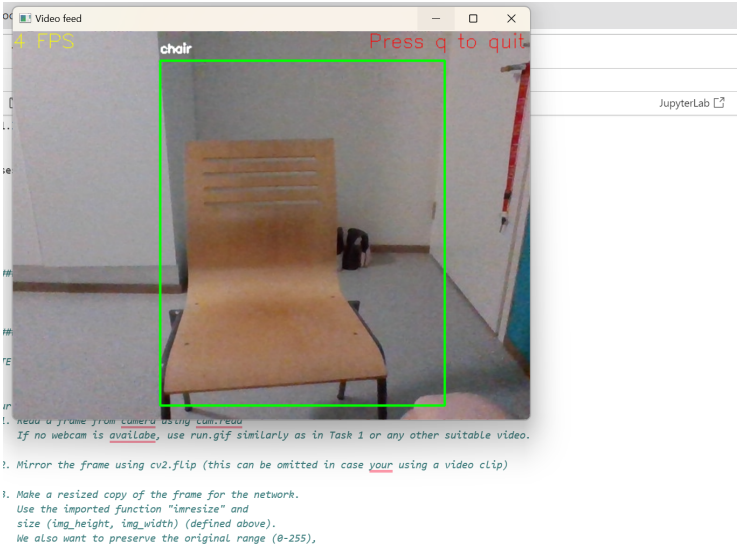
In general, training set is used to train the model while validation set is used to fine-tune the model. We don't use the test set for fine-tuning, otherwise, the model will be biased towards the test set.

b) There are 7 convolutional blocks. Each block is built from:
Convolutional Layer (Conv2D)
Batch Normalization Layer (BatchNormalization)
Activation Layer (ELU - Exponential Linear Unit)
Pooling Layer (MaxPooling2D)

c) Two attributes of the loss function
Localization loss and confidence loss
- Localization loss measures the accuracy of predicting the bounding box coordinates for each object in the image.
- Confidence loss evaluates the model's ability to predict the presence or absence of objects in each anchor box.

3


Video feed
FPS
Press q to quit
person
bus
Search
ENG
17:57
27/01/2024


Video feed
4 FPS
chair
Press q to quit
JupyterLab
If no webcam is availabe, use run.gif similarly as in Task 1 or any other suitable video.
2. Mirror the frame using cv2.flip (this can be omitted in case your using a video clip)
3. Make a resized copy of the frame for the network.
Use the imported function "imresize" and
size (img_height, img_width) (defined above).
We also want to preserve the original range (0-255),