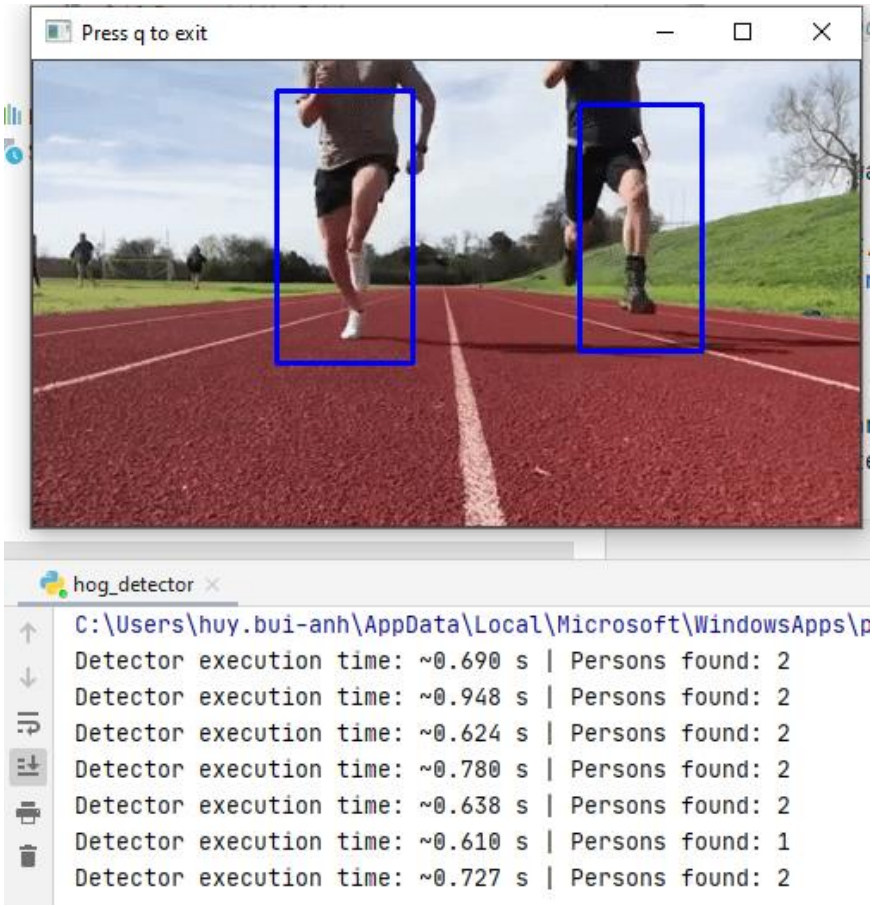
Computer Vision – Week 3

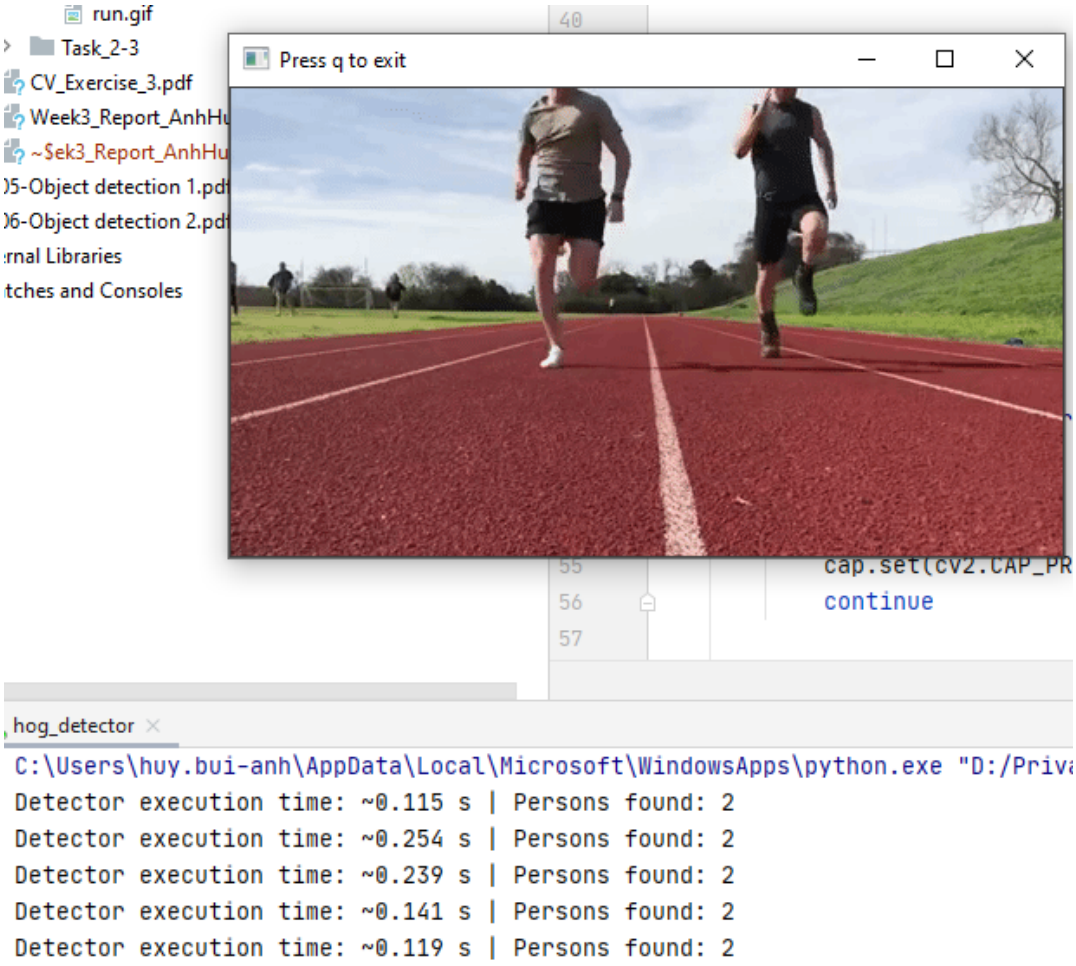
Anh Huy Bui – 293257

1. With:
    - scale = 1.07
    - winStride = (3, 3)
    - padding = (8, 8)
    - hitThreshold = 0.0



With:

- scale = 1.06
- winStride = (4, 4)
- padding = (8, 8)
- hitThreshold = 0.0



Faster but at frame 6, no person is found compares with above with 1 person is found

⇨ Less accurate

2.

a. The target objects are classified into different categories called “class_id” and information for each target is described with location of 4 pixels that form a boundary box of that target object.
   Training set: a set of data samples used to train the model.
   Validation set: data samples in this set are used to provide an unbiased evaluation of the trained model in different condition (while tuning model hyperparamters).
   Test set: data samples in this set are also used to evaluate the final-tuned model.

b. There are 7 convolutional "blocks". Each block consists of many layers which execute convolution of the output of previous layer with kernels. All of these blocks contain 1 stage of max-pooling on the output:
   Block 1: output of lambda layer * 32 kernels with size 5x5
   Block 2: output of 1st block * 48 kernels with size 3x3
   Block 3: output of 2nd block * 64 kernels with size 3x3
   Block 4: output of 3rd block * 64 kernels with size 3x3
   Block 5: output of 4th block * 48 kernels with size 3x3
   Block 6: output of 5th block * 48 kernels with size 3x3
   Block 7: output of 6th block * 32 kernels with size 3x3

c. SSD loss function consists of 2 paritial losses: localization loss and confidence loss.

   Localization loss defines bounding box regression (calculate differences between predicted box and ground truth box). Smooth L1 is mentioned in the publication as a method for localization loss, because it combines advantages of both L1 and L2-loss.

   “The confidence loss is the softmax loss over multiple classes confidences.”

   - arXiv:1512.02325v5 [cs.CV] 29 Dec 2016 - SSD

3.


Image
person 1.0
person 0.87
car 0.99
person 0.9
person 0.92
3 FPS
Press q to quit

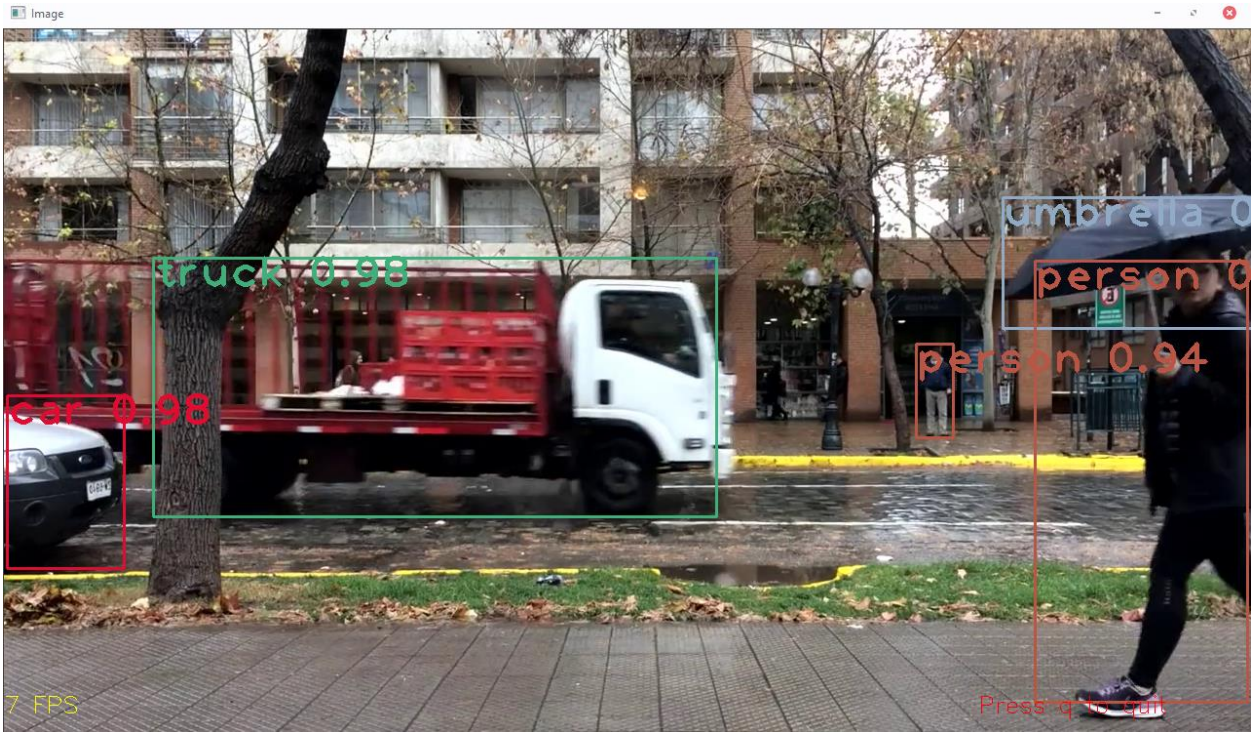


Image
truck 0.98
car 0.98
umbrella 0
person 0
person 0.94
7 FPS
Press q to quit