1. I predicted all classes on the given video (i.e. ignore classes[i] == 1) and it output many boxes. Figs are shown below.

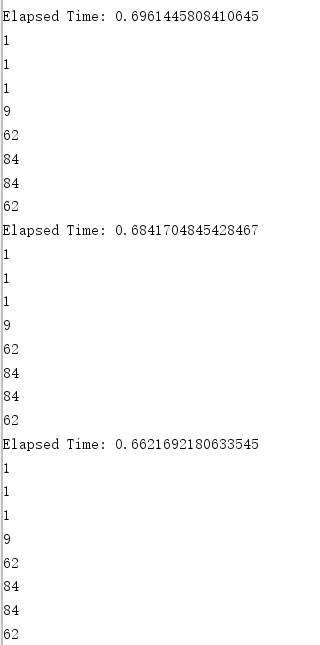


Fig 1



Fig 2

Sadly, I didn’t find any statement that can label what class it is on the output box as well as print the predicting probability. So I added ‘print(classes[i])’ statement to see the classes.



Related to Fig1

And we know:

class1: person; class9: boat; class62: chair; class84: book;

Though the result seems not precise enough for non-human classes, I think it is forgivable because other classes are too similar, like books and square well covers, oil tanks and boats.

2. I read some papers and related implementation about counting/tracking people.

I knew the history and major face recognition methods, like ASM、AAM、CPR before CNN, and DCNN、TCDCN、MYCNN etc. But these are not based on CNN (the network are changed), like the model I chose, so I don’t think they can merge into the current project. What’s more, these methods focus on face recognition rather than verify and count people.

Then I read some paper about people tracking and re-identification.

One paper I read is: <http://openaccess.thecvf.com/content_cvpr_2017/papers/Tang_Multiple_People_Tracking_CVPR_2017_paper.pdf>

Here’s another two: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8486568>

[REAL-TIME MULTIPLE PEOPLE TRACKING WITH DEEPLY LEARNED CANDIDATE.pdf](REAL-TIME%20MULTIPLE%20PEOPLE%20TRACKING%20WITH%20DEEPLY%20LEARNED%20CANDIDATE.pdf)

I’m not fully understand their theory and operating circumstance, but the idea may worth trying.

And here’s an article about different ideas of multi-target tracking in Chinese: <https://blog.csdn.net/sigai_csdn/article/details/83303643>

It summarizes several good paper of multi-target tracking.

Till now I have several questions：

1. What’s the final goal of the project? If I remember right the goal is to track people’s movement and predict violence, then I think we won’t actually use human-detect models because tracking models can detect human as well. Is that right?
2. When counting people in the whole video, is it suitable to see someone as two if he appears then disappears, and re-appears then re-disappears? This can skip people re-identification and still useful to track people’s movement and predict violence.
3. The test of many paper I read about people tracking and re-identification is from a positive perspective to human body and I think they may not processing well in wild videos.