May 15, 2020

Author: **Joram Bakekolo**

Report Computer vision Lab 3

Abstract

In this lab we built a simple object tracker. The tracker detected objects on all frames of a video and linked the predictions from one frame to the text. We explored algorithms to link our predictions across time frames. Indeed, tracking involves involves not only detecting objects in each frames t=0, but also linking them across time where the target is in frame $t > 0$.

1 Part A: Download Data

For the implementation, we used colab notebooks to train detectron2 models. That was essential for it colab GPU compared to the personal laptops which only have CPU. We strated by downloading a small video clip of 41 frames source), and loaded the first frame of the video. We created a detectron2 configuration and a detectron2 DefaultPredictor to run inference on our images. Then we visualized the predictions from a random set of frames to make sure things look correct.

(a) Predictions from random set (b) Prediction from random set

2 Part B: Tracking Objects in Pairs of Frames

In this part we tried to track objects in time based using the detectron 2 predictions on each frame of the video. To do it we defined two frames and we defined a function to see the matching score between the objects in the two images. The function helped to link two predictions if they belong to the same object category and if their bounding box overlap is high. We visualized it and obtained the figures below

(a) Tracked frame 0 and 1 (b) Tracked frame 20 and 21

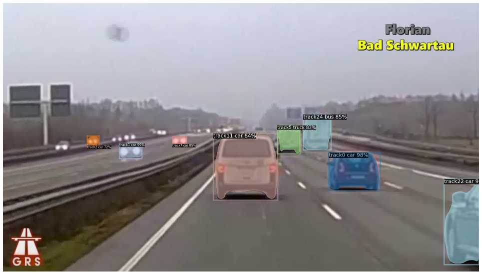
(a) Tracked frame 39 and 40

3 Part C: Tracking Objects in Videos

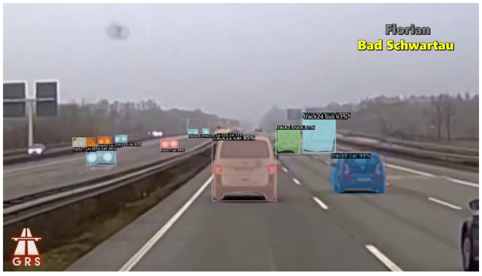
Now we want to track for longer horizons such that we can see at each steps (dt> 1). To do it we extended the technique used in part B consecutively. We picked 10-frame clip from the video, and we visualized the 10-frame track. We used colorcoding to indicate the predictions which belong to the same track.

(a) Tracke of 10

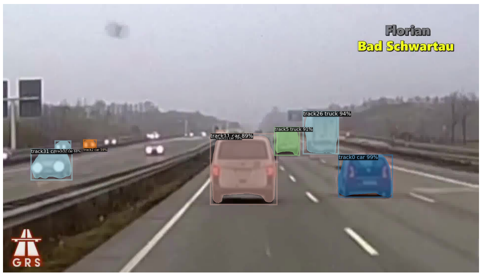
(b) Tracke of 10



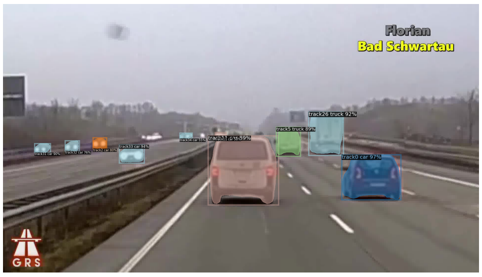
(a) Tracke of 10



(b) Tracke of 10



(a) Tracke of 10



(b) Tracke of 10

We observed that some time the model does not tracke the objects. For example when the car moved to the otherside, some time the model did not detected the car, some time the car change to be a tracker. In case of two frame at a time when the object is moving fast, the model is not able to track the object. But in case of ten frames, the model is able to capture the object even if it is moving fast, the model is able to detect the next object if it missed the first.

4 Conclusion

In conclusion, the model was consistant in tracking objects some time butalso made mistake caused by the mouvement in time. The model has some problems because we know the tracker is not the model but the Mark R-CNN which is doing object detection and Mark R-CNN is not 100 percent accurates.