In this exercise, you need to track the detections made by CenterNet using SORT or DeepSORT.

We are going to use a video from The VIRAT Video Dataset (use one in the Ground Dataset, not Aerial Dataset). This dataset contains videos in the surveillance domains. You can choose any of them. One example is VIRAT_S_010204_05_000856_000890.
We are again going to use CenterNet but this time one there has been pre-trained on the COCO dataset. The model will predict the COCO classes. You can choose to track one or more classes. You can also do tracking on the human poses with the model.

The tracked bounding boxes should then be visualized on the input image. For this, you can use the VisTrack class.

I have again made a Colab notebook ready. To use GPU (or TPU), go to Runtime > Change runtime type and under Hardware accelerator, choose GPU or TPU.