REPORT FOR ASSIGNMENT 2

Muntashir Bin Solaiman

Student ID - 22021468

Learning how to detect objects in a video was the objective. A pretrained Mask R-CNN model was used. The objects were first detected. They were tracked across the video. The results were put in a video.

The objects were first detected. For a detection, its box, label and score were captured. The box contained the object's coordinates. How confident Mask R-CNN was with the classification was represented by the score. Score lower than 0.5 was not considered. The detection was run through the entire video. It was done frame by frame. The 3 things that were captured were cumulated in lists of that type. The lists were saved in files. They were filtered_boxes.pt, filtered_labels.pt and filtered_scores.pt. The methods used were filter_frame_detections(), filter_all_frames_detections() and save_all_detections().

After this the objects were tracked. When an object was detected, a track was initialized. The track contained the objects label, traces of its coordinates throughout the video and an id. Whether a new detection was of an existing track or not usually was the problem. Different objects of the same class could be in a video. The assumption was same object will not move large distances in two frames. Hence the Euclidean distance between the box's centers was calculated. In tracks, classes that matched the detected ones were narrowed. Their box centers were calculated. Their centers were subtracted. The objects were matched with two conditions. Their distances were the lowest and did not exceed the maximum distance. The methods used were get_box_center(), calculate_center_distance() and track_objects().

The output video was created. The bounding boxes and labels were drawn. It was done on one frame at a time. The frame was put into the output video. The methods used were draw_tracking_results() and create_tracking_video().

The video had bounding boxes overlayed. The detected objects had their class number. The identified classes were 1, 27, 51, 53, 63, 72, 77. They were person, backpack, bowl, apple, couch, tv and cell phone. There were two couches. They had different color bounding boxes. This model identified them separately. Laptop was misidentified. It was identified as a bowl. A limitation was found. The model had trouble distinguishing objects of different classes if they were close to each other. This happened because of how the bowl was placed relative to the laptop. The video was played for 5 seconds. The bounding box for the couch on the left disappeared. The bounding box on the laptop disappeared. My hand was raised. It did not cause a problem. As the hand was raised the bounding box of the phone moved to the left. In the 11th second the bounding box of the phone disappeared. The disappearing boxes could be fixed by lowering the confidence score.

The model was very accurate. It only misidentified one object. This could have been fixed by placing the laptop and the bowl separately. To fix disappearing bounding boxes, the confidence threshold was lowered. This did not make any changes.