Hellometer Excercise: Alberto Rivera

Overview

This exercise was broken down into 3 pieces, Object Detection, Object Tracking, and Analysis. A video showing the output of the entire process can be seen on this video here. All the code is also available in this github repository.

Screenshot from video overview video

Object Detection

For the object detection I started with a SSD Mobilenet TensorFlow model that was pretrained on the COCO dataset. The model was originally trained to classify 90 classes, which include ‘person’. Since we are only concerned with the person class we can improve performance of the model by fine tuning the model to only detect people.

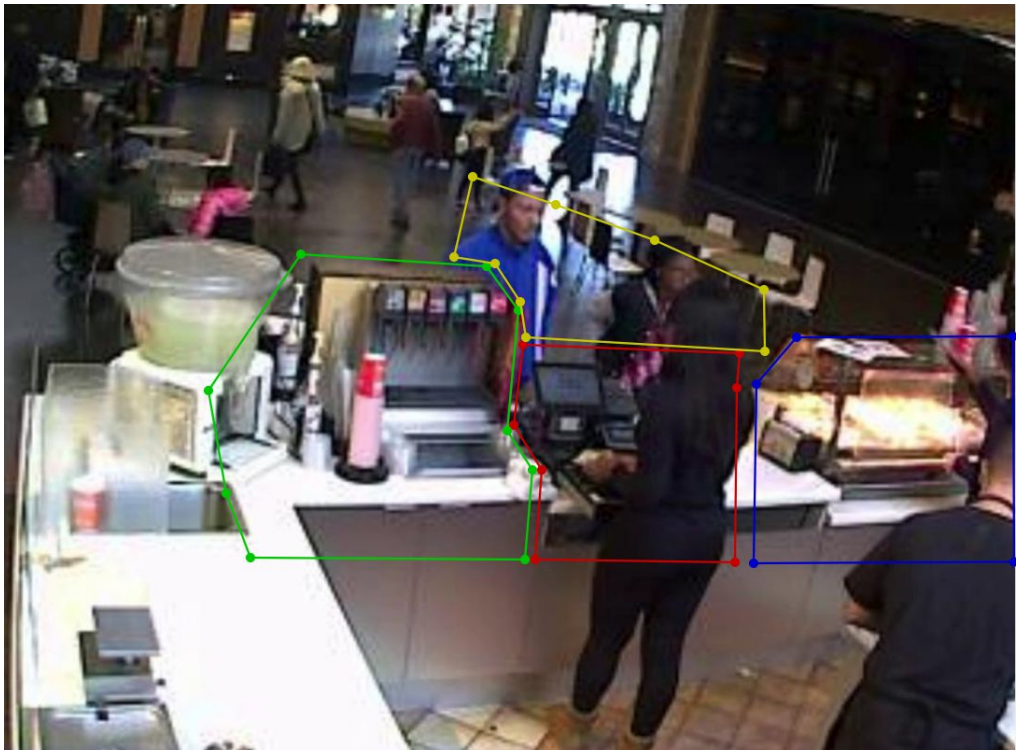
I hand labeled bounding boxes around people on a few frames from the **food counter** video and used them to re-train the model. This improved the accuracy of the object detection model greatly. Once the detections were fairly accurate we can build the rest of our system around them.

Object Tracking

With decent object detection our next task is to match detections to the same person. For this process we keep track of all the object detections on the previous frame. For every detection in the current frame we use a heuristic to match the new detection to a detection in the previous frame. If a new detection has no match, we assume that it is a person we haven’t seen before. Good object tracking is crucial to determine customer wait times, since we have to detect that the same person has been waiting in line.

Area Usage Analysis and Customer Wait Times

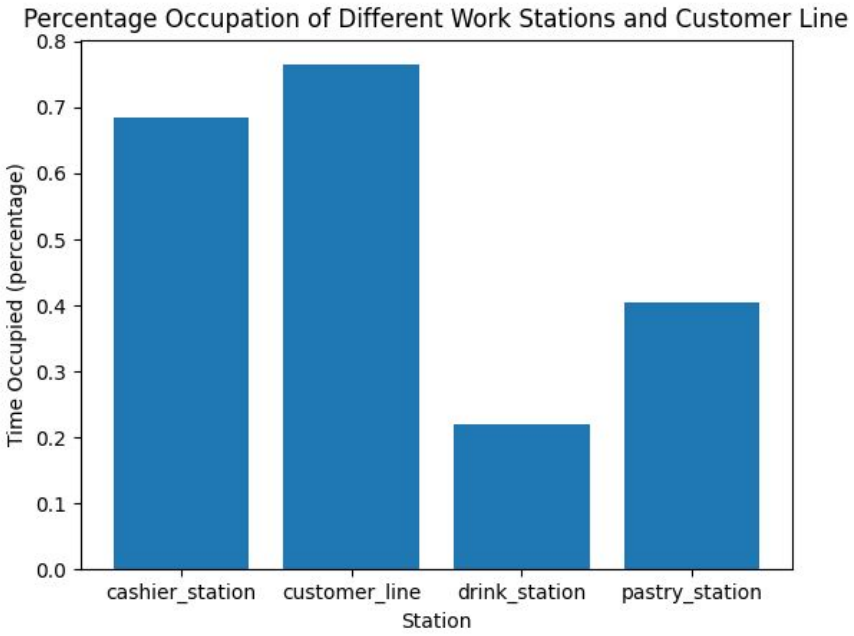
With decent detection and tracking out of the way, we can now focus on analysing how the different employee stations are being utilized as well as calculating customer wait times. To achieve this we first hand label the Areas of Interest (AOI). In the image shows the areas we determined:



Drinks station: green, Cashier station: red, Pastry station: blue,
Customer line: yellow

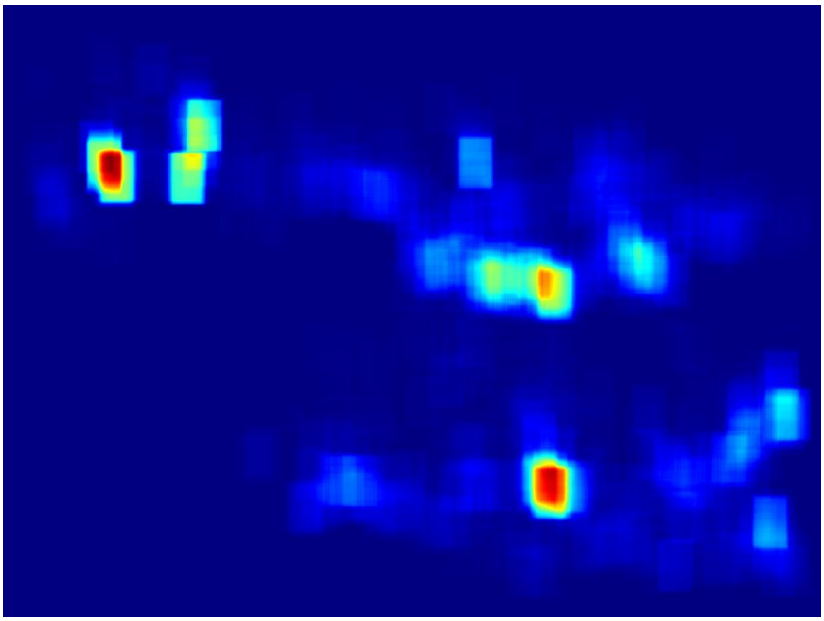
With the AOIs in place we can determine which people are in a given area at any given time. This is done by calculating the center of each person's bounding box and if that center is within one of the AOI polygons we consider that area to be in use. For this particular video we assume that the frame rate is 1 FPS, so we add one second of usage to our calculation for each frame that at least one person is in that area.

After running the video we can visualize our results:



The figure above shows the percentage of time that an area was occupied. The areas with the heaviest usage were the customer line with close to 80% and the cashier station ~70% usage. We can already tell from this graph that for ~10% a customer was in line there was no employee behind the register. We also see that a significant amount of time is spent at the pastry station.

Using a similar technique we can determine how much time a customer waited in line.



The final analysis done for this exercise was to create a heat map. This was done by creating a black image and adding +1 to the pixels around the center of a detected person. In the image above we can see 4 main bright spots. The two upper left bright spots actually correspond to one of the tables in the background that had a couple of different people occupy it for the duration of the video. The centermost bright spot is the area of waiting customers. The spot under that is the cashier station, which was the most occupied by the employees.

Improvements and Future Work

Most of this work was done quickly as a proof of concept and can be improved greatly.
The main area for improvement is the object tracking. Fine tuning of the tracking heuristic parameters could help increase accuracy by reducing object id switching, which is the main problem currently. Part of this problem is also caused by occlusion due to the camera location. Opting for a neural network approach for tracking instead of a traditional Computer Vision approach might also improve performance.