

UCSD Machine Learning Bootcamp Capstone Proposal

I am proposing a SaaS as my capstone project. The SaaS will allow businesses like grocery stores, supermarkets and any other business with pedestrian foot traffic track customers within the business premises. We are using a variety of data sets including the virat data set to train our model. These datasets are all open sourced. To solve this problem we will use one of the following deep learning algorithms Yolov3, Faster R-CNN, Single Shot Detector (SSD) to detect persons in combination with deep sort to track persons..

This is a supervised learning problem since we know what objects we want i.e. people and our datasets provide bounding boxes for them to train the model. This is also a classification problem since we are detecting objects in our case people from all of the other objects in the video. We are predicting bounding boxes and using images as our predictors.

Our final deliverable is a web application that will allow its user to submit mp4 videos. Our application will then output an annotated video with bounding boxes around the persons detected with the video. Our application will also allow its users to download both the annotated video and a CSV file containing bounding boxes coordinates throughout the video.

Due to the nature of our problem we will need significant GPU processing power in order to provide close to real time/ acceptable delay between uploads and results. Furthermore, we will require 16-32 GB of RAM and two CPUs to ensure decent performance for our applications at low traffic/ upload volumes.