Abstract: Self-driving will be the leading way of transportation in the future which can make fewer mistakes and save time. It also can be used in many areas, such as the virtual reality games, intelligent wheelchair, and auto-drive sightseeing cars. The technology of self-driving will bring big convenience and huge pleasure for human. The combination of using data gathering from camera, lidar, radar, and HD map will give multi-ensure when the decision makes, the high accuracy of multiple targets detection in the image is one of the critical parts of self-driving decision making. The main work I want to do in this project is detect multi targets including biker, car, pedestrian, traffic Light, traffic Light-Green, traffic Light-Green Left, traffic Light-Red, traffic Light-Red Left, traffic Light-Yellow, traffic Light-Yellow Left, and truck in images. Since multiple objects are detected all at once, YOLO (you only look once) will be used to do the detection in this project. YOLO is a real-time object detection system which can identify multiple specific objects in videos, live feeds, or images. It has versions from 1 to 5 now. I plan to use YOLOv3, because it is fast, accurate stable and fully developed. Different from other state-of-the-art methods in object detection (e.g., R-CNN, fast R-CNN, Faster R-CNN) which runs regional proposals, YOLO only runs the network at once to do the inference which makes it faster than the R-CNN-based methods. The pre-trained model which trained on big datasets, such as COCO dataset, VOC dataset will be used as an initialization of the training. The dataset I will use in the project comes from Udacity Self Driving Car Dataset which consists of 15000 images with 512\*512 pixels. It has several different data structure formats for YOLO, RetinaNet, etc, which can be download directly from the website. IOU (Intersection over union) and MAP (mean average perception) will be used as the evaluation criteria. IOU is commonly used as an evaluation for YOLO, and MAP is always used to evaluate the performance of object detection.