Abstract

With the rapid development of automobile industry, Internet technology, big data and other technologies, automatic driving technology combining automobile industry with computer science emerges as the times require. The current automatic driving vehicles follow the process of “perception-decision-control” .The basis of all downstream tasks is perception module, such as object detection, instance segmentation, object tracking and so on, which provide key information for the motion decision and path planning of unmanned vehicles. The application of the deep learning method to solve the perception problem can produce high-precision calculation results, and the high-precision model generally takes up a lot of memory and computation. But the computing devices of the vehicle platform have low power consumption and computing power. At the same time, single sensor perception has defects of field of view or information representation, multi-sensory perception scheme needs to solve problems such as perception conflict resolution, representation unification and delay synchronization.In order to solve above problems, this research is committed to the research of lightweight multi-object tracking (mot) model based on different data sources, aiming to propose real-time, accurate and robust multi object tracking method, and improve the real-time and reliability of perception results in the process of automatic driving;

1) The vision-based mot model is mostly detection-based tracking. In order to solve the problem of noise in the bounding boxes, a high-precision two-stage network model named Fast-TrackRCNN and a high-efficiency single-stage network model named PolarTracker were proposed respectively.Fast-Trackrcnn combines the Positional Normalization (PONO) feature with a lightweight CNN backbone for refining object segmentation. The bi-directional Long Short Term Memory (Bi-LSTM) module is then used to further enhance the association information of inter-frame targets. Compared with the original TrackRCNN, our model is two times faster than Track RCNN with slight drop 2% on sMOTSA, and is more suitable for deployment of vehicular low-power edge computing equipment.Based on the single-stage structure, PolarTracker replaces the traditional mask prediction by regression of a group of fixed edge points in polar coordinate system, which

can greatly optimize the computational complexity and regression difficulty of the mask. The model achieves 630.91FPS on 2080Ti and 127.06FPS on AGX after using TensorRT.

2) Through the fusion of spatial information of point cloud and texture information of the image, an efficient multi-sensor fusion based object tracking model PV-Enconet-Track was proposed. Firstly, the point cloud is projected onto the image based on the external parameter, and the model eliminates a large number of invalid point clouds through a ground filtering algorithm, which reduces the amount of data by about 50% and increases the speed by about 42%. Then, a point encoding module is designed based on the Neighbor Sampling with Background (NSB)module, the Spatial and Texture Encoding (STE) module and attentive pooling layers, and combined with the lightweight voxel encoding network, generate features of each colored point, and then predicts the position, heading and class of the objects. Our model is 7 FPS faster on the 2080Ti than the classic PointPillar network,and 2 %, 18 %, and 10 % better than PointPillar in vehicle, pedestrian, and cyclist detection precision. Finally, the mot task is accomplished by fusing the results of 3D detection and 2D detection . The tracking performance of PV-Enconet-track is about 5% higher than Fast-TrackRCNN and 12% higher than PolarTracker.

3) In order to verify the effectiveness of the above algorithms in actual scenarios, mot dataset is collected and labeled through local vehicle platforms. In order to speed up data annotation, a semi-automatic data annotation method is designed to generate 2D and 3D rough annotation results. The dataset consists of 3500 frames data for local training and evaluation. Through testing on local dataset and vehicle computing platform, it is verified that this scheme balances the computing power of vehicle computing equipment and the depth of the deep learning model, and completes the mot task efficiently and accurately.

Key words: Object Detection, Object Tracking, Vehicle Platforms, Autonomous Driving, Deep Learning