Attention-based Temporal Frustum PointNet for 3D Object Detection

Technical University of Munich

Faculty of Informatics

Chair of Robotics, Artificial Intelligence and Real-time Systems

Supervisor:
Prof. Dr.-Ing. Alois Knoll

Advisor:
M.Sc. Emeç Erçelik

Research project:
-

Type:
Master's Thesis

Research area:
3D Object Detection, Autonomous Driving

Programming language:
Python

Required skills:
Neural networks, Tensorflow

Language:
English

Date of submission:
March 25, 2021

For more information please contact us:

Phone: +49.89.289.18130

E-Mail: emec.ercelik@tum.de

Internet: https://www.in.tum.de/i06/people/emec ercelik-msc/

Background

Object detection is an important part of the perception mechanisms of the autonomous cars. The aim of object detection is to localize the objects surrounding the ego-vehicle and identify the type of surrounding objects. With the increasing availability of public datasets that include LiDAR points [1,7,8], object detection is mainly conducted in 3D space for autonomous driving function. Even though the quality of sensors increase day-by-day, there are still challenges in front of having perfect detection results such as occlusion due to dynamic objects and sparsity of depth information provided with LiDAR data. These problems can be alleviated fusing sequential information in time [5,6]. Temp-Frustum Net [9] is one of the recently proposed methods that tackle this problem using temporal fusion.

Description

In this thesis, the student will be working on the Temp-Frustum Net [9] to improve its temporal fusion module using an attention mechanism. Temp-Frustum Net is an extended version of Frustum PointNet architecture [2]. The Frustum PointNet is a seminal work for PointNet-based 3D object detection methods. For 3D detection, the 2D bounding boxes are predicted on images and points in the frustum of each 2D bounding box are extracted. The extracted points per frustum volume are further segmented to find only the points that are of the object of interest. Afterwards, 3D Amodal Box Estimation PointNet generates global feature of the object of interest and finally 3D bounding boxes are predicted using the generated global features. Temp-Frustum Net extends this architecture by applying a temporal fusion to the global features of the objects in successive frames. However, the proposed module relies on the ground-truth 2D bounding boxes and tracking IDs as a proof of concept, whose quality also affects the 3D detection performance. This weakness will be also addressed in this thesis using off-the-shelf 2D detector and tracker architectures. Additionally, the temporal fusion that is realized by GRUs will be supported by attention mechanisms.

Tasks

This student project consists of the following tasks:

- Reviewing 2D object detection and tracking architectures [1,2] (2-3 weeks)
- Selecting and running one of the architectures that can provide a high recall and small number of ID Switches on KITTI tracking dataset (1-2 weeks)
- Obtaining results on Temp-Frustum Net using ground-truths and also using the tracking results. (2-3 weeks)
- Reviewing temporal attention modules to integrate into Temp-Frustum Net [3,4] (2 weeks)
- Benchmarking the currently integrated attention module (2 weeks)
- Extending the Temp-Frustum Net with the selected temporal attention module (3 weeks)
- Benchmarking the extended Temp-Frustum Net using the ground-truth tracking information and also the estimated tracking information (6 weeks)
- Reporting, preparation of a new paper out of results and README for the modifications (3 weeks)

References

1. Geiger, Andreas, Philip Lenz, and Raquel Urtasun. "Are we ready for autonomous driving? the kitti vision benchmark suite." 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012.

2. Qi, Charles R., et al. "Frustum pointnets for 3d object detection from rgb-d data." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
3. Wang, Xiaolong, et al. "Non-local neural networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
4. https://paperswithcode.com/methods/category/attention-mechanisms
5. Yin, Junbo, et al. "Lidar-based online 3d video object detection with graph-based message passing and spatiotemporal transformer attention." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
6. Yuan, Zhenxun, et al. "Temporal-Channel Transformer for 3D Lidar-Based Video Object Detection in Autonomous Driving." arXiv preprint arXiv:2011.13628 (2020).
7. Caesar, Holger, et al. "nuscenes: A multimodal dataset for autonomous driving." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
8. Sun, Pei, et al. "Scalability in perception for autonomous driving: Waymo open dataset." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
9. Erçelik, Emeç, et al. "Temp-Frustum Net: 3D Object Detection with Temporal Fusion." 2021 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2021 (submitted)

Technical University of Munich

Faculty of Informatics

Chair of Robotics, Artificial Intelligence and Real-time Systems