Seminar SS2020 - Visual Feature Learning in Autonomous Driving

Erçelik, Emec

April 2020

1 Introduction

In this seminar, we will be focusing on object detection and tracking problems. Object detection tries to localize objects with their classes on 2D plane (frontview or bird’s eye view) or in 3D space. This provides the distance of the ego-vehicle to objects in the environment as well as their shapes, which is useful for planning algorithms. Tracking is another aspect of the localization, in which the position of specific objects are determined in subsequent frames assigning a tracking ID to these objects. Therefore, these two tasks for autonomous driving are very related to each other and two very basic problems for autonomous driving function.

You will be providing a literature review with one of the topics in Section 2 in a group. The groups will be formed with minimum 2 people and maximum 3 people. Therefore, only some of the topics that receive most attention will be assigned to the groups. I will ask you in the introduction session to vote three of the topics according to you or your group would like to review most. You can send you choices via email on the same day. If you can’t form a group, the people assigned to the same topic will be a group and will be working on the same report. The report is a 6-8 page document of your findings in IEEE format.

Under the topics I provide a few references that I found useful for starting to review these topics. You don’t have to include all the provided references in your research report. Before making a decision on a topic, please read my explanation on this topic and abstract, introduction, and conclusion sections of the provided references to have a broad understanding of the topic. It is better to start reading a paper with the abstract and introduction than directly going into details of a paper. After the topic assignment, you can continue your research with different methods. Here are some suggestions:

- Checking cited papers in the references section of the provided papers (To find similar studies)

- Checking studies that cited the provided papers (To find more recent studies, for example using Google Scholar)
• Checking leaderboards of the public datasets

• Checking the conferences related to the topic, i.e. ICCV, CVPR, IV, ITSC

2 Topics

• **3D Object detection using different data modalities:** (max 3 students) For the autonomous driving function in real-world, one of the key properties is the ability of 3D object detection, which provides precise locations and sizes of objects in the environment in order to do path planning for future positions of the ego-vehicle. Due to the weaknesses of several sensor types, it is argued to be better when data from several sensors are fused to compensate each other. The fusion of sensors is useful when a few of the sensors fail to provide proper information. In this seminar topic, you are asked to review such challenges brought by the weaknesses of the sensors, the sensor fusion methods for 3D object detection using different modalities. Additionally, you will be providing your own ideas as solutions to these challenges with evidences from the literature. You can find some of the sensor types you can include below, but you are not limited to these.

  – LiDAR
  – Radar
  – RGB or gray-scale camera images
  – Multi-camera images
  – GPS / maps
  – Review on sensor fusion types

References: [5, 11, 21, 12]

• **Comparison of only LiDAR-based and fusion-based 3D object detection methods:** (max 2 students) 3D object detection mostly relies on LiDAR sensors, which yield highly precise depth information of objects in 3D space. Even though it has its own advantages, LiDAR information is usually sparse comparing to camera images and doesn’t work properly in different weather conditions as well as radar sensors. With these disadvantages in mind, the state-of-the-art methods evaluated on public datasets currently utilize only LiDAR points. In this seminar topic, you are asked to compare methods using only LiDAR information for detection with the sensor fusion methods for 3D object detection from different perspectives by reviewing the literature. In the end, you will be also providing your own ideas about which method to utilize for better detections with the evidence from literature. Some of the points of view might include:

  – Advantages of one over each other
– Safety concerns
– Accuracy of methods
– Possible saturation in results with the future improvements in the architectures
– Extendability of architectures based on sensors
– Upcoming improvements in the sensors

References: [12, 11]

• **Challenges in video object detection and tracking for 2D and 3D bounding boxes: (max 2 students)** Detecting objects might be challenging due to the low quality of sensor outputs at the time of sampling (reflection etc.) and occluded objects behind others (parked cars, not fully seen behind the first few cars). These problems can be relieved by using time-series data such as videos, or sequences of LiDAR point clouds. In this topic, you are asked to review recent studies about object tracking for 2D and 3D bounding boxes and provide current challenges of these problems with the solutions suggested in the literature. Additionally, you are asked to provide your own solutions for these challenges considering the experiments in the literature.

References: [1, 20, 15]

• **Convolutional recurrent neural networks (RNNs) for object detection and tracking: (max 2 students)** RNNs are the most commonly used architecture types to keep track of important temporal features to improve accuracy for the task using time-series data. In object detection and tracking domains, using RNNs might get tricky due to multi-objects in one scene. In this case, the matching problem arises, which aims at aligning features from the same objects in time. However, convolutional recurrent layers can be considered as a solution to this problem, which processes feature maps of the entire scene instead of processing features of objects separately in time. In this topic, you are asked to review the usage of convolutional recurrent layers in the domain of object detection and tracking. In the end, you will be also providing your own ideas for the usage of these layers, usefulness depending on data type, and the feature map type. Some of the aspects might be, but not limited to the ones below.

  – Advantages over linear RNNs
  – Comparison on cell types (e.g. LSTMs vs. GRUs)
  – Accuracies on different data types (camera images, radar data, LiDAR data)
  – Comparison on the efficacy for object detection and for tracking

References: [25, 14, 19, 15]
• **Semi-supervised, self-supervised, and unsupervised learning methods for object detection:** (max 3 students) Supervised learning is the most commonly used learning method for training object detection networks. However, the more complex data types are, the more difficult and expensive become the labelling efforts. Therefore, there have been an increasing interest in training networks with small labeled datasets aiming to reach similar accuracies with the supervised learning. In this topic, you are asked to review such methods for object detection especially in 3D space. Some aspects to be considered while reviewing might be the following:
  - Applicability to different data types
  - Comparison of results of these methods with results of methods based on supervised learning
  - Differences in loss functions used for these training schemes

References: [22, 24, 17, 9, 18, 10]

• **Transfer learning for 3D object detection:** (max 3 students) Deep neural networks are mostly known to be data hungry, which also applies to the 3D object detection, for which most studies rely on large networks. Even though collecting data from real-world environments is relatively easy, synchronizing and calibrating multiple sensors, labeling the collected data for the required tasks, and ensuring the diversity in the object classes, weather conditions, illumination, road types are difficult, time-consuming, and expensive. Therefore, there have been attempts to transfer the information learned from the available datasets to the new datasets to decrease efforts on preparing new datasets. Additionally, there have been virtual datasets and simulation environments as a solution to this problem. In this topic, you are asked to review performances of studies using the available real-world and virtual datasets for transfer learning in 3D object detection as well as available simulators. You will be also providing what kind of sensors and diversity these datasets and simulators have to offer in addition to the your ideas for improving such knowledge transfer with the evidence from literature.

References: [17, 23, 9]

### 3 Other Useful Resources

- **Datasets:** KITTI [8], nuScenes [4]
- **Simulation environment:** CARLA [6], GTA V [16]
- **Metrics:** Detection [13], Tracking [2]. Keywords: Average Precision, Intersection over Union, MOTA, MOTP
- **Survey:** [7]
References


