2020 IEEE Intelligent Transportation Systems Conference (ITSC)
Rhodes, Greece. September 20-23, 2020 (Virtual)

The rounD Dataset: A Drone Dataset of Road User Trajectories at Roundabouts in Germany

Robert Krajewski[1], Tobias Moers[2], Julian Bock[2], Lennart Vater[1] and Lutz Eckstein[1]

Fig. 1: Exemplary road user trajectories at one of the roundabouts in the rounD dataset. By applying deep neural networks to traffic recordings taken by a drone, an accurate track for each road user is derived from aerial imagery.

***Abstract*— The development and validation of automated vehicles involves a large number of challenges to be overcome. Due to the high complexity, many classic approaches quickly reach their limits and data-driven methods become necessary. This creates an unavoidable need for trajectory datasets of road users in all relevant traffic scenarios. As these trajectories should include naturalistic and diverse behavior, they have to be recorded in public traffic. Roundabouts are particularly interesting because of the density of interaction between road users, which must be considered by an automated vehicle for behavior planning. We present a new dataset of road user trajectories at roundabouts in Germany. Using a camera-equipped drone, traffic at a total of three different roundabouts in Germany was recorded. The tracks consisting of positions, headings, speeds, accelerations and classes of objects were extracted from recorded videos using deep neural networks. The dataset contains a total of six hours of recordings with more than 13 746 road users including cars, vans, trucks, buses, pedestrians, bicycles and motorcycles. In order to make the dataset as accessible as possible for tasks like scenario classification, road user behavior prediction or driver modeling, we provide source code for parsing and visualizing the dataset as well as maps of the recording sites. For non-commercial public research, the dataset is available free of charge at https://www.round-dataset.com**

[1]The authors are with the research department Vehicle Intelligence & Automated Driving, Institute for Automotive Engineering, RWTH Aachen University, 52074 Aachen, Germany {krajewski, vater, eckstein}@ika.rwth-aachen.de

[2]The authors are with fka GmbH, 52074 Aachen, Germany {tobias.moers, julian.bock}@fka.de

I. Introduction

Future traffic will be shaped by automated vehicles. While many companies and researchers are trying to automate road traffic, it is becoming clear that developing and validating these systems is a highly sophisticated problem. Due to their complexity, many challenges such as modelling natural driving behavior [1], predicting the intentions and movements of surrounding road users [2] or classifying possible driving situations [3] can no longer be solved by classical methods. An automated vehicle can be in a vast number of possible scenarios, which include not only variations of the static infrastructure such as road layouts but also dynamic elements such as other road users and their behavior. Due to this diversity, state-of-the-art-solutions often rely on data-based methods rather than on expert knowledge. As a basis for these methods, a large trajectory database is needed that contains the naturalistic and diverse behavior of road users on public roads.

The required data cannot be generated with current methods in simulations or by systematic tests on a test track, because the recorded behavior would be biased and would not include corner cases that are likely to occur in real world conditions. However, data collected on public roads includes typical misconduct such as speeding or negligent driving if people are not aware that they are being recorded. As this naturally occurring behavior has to be accounted for in

978-1-7281-4148-0/20/$31.00 ©2020 IEEE

automated driving, data of this kind must be collected by an efficient method to provide a sufficiently large amount for the development and validation of automated driving systems. In recent years, different approaches have been developed that allow to create such datasets. In addition to equipping vehicles with sophisticated measurement sensors such as one or more lidars [4], statically positioned infrastructure sensors are also used to detect road users [5]. However, the aerial perspective has proven to be a very advantageous option compared to the other methods, as it is unsurpassed in terms of absence of influence on traffic, perspective, flexibility and data quality [6].

Particularly interesting are situations in which a system has to interact with surrounding road users. In order to handle these safely, an automated driving system must correctly classify the driving situation. This includes not only precisely sensing surrounding road users but also reliably predicting their movements. Only if these requirements are met, a system will be capable of planning a trajectory that is safe and comfortable for all road users.

Traffic scenarios at unsignalized roundabouts typically include a high degree of interaction between different types of road users. In comparison to highways or structured environments, roundabouts are more complex due to the variety in maneuvers and need for negotiation between road users. When entering a roundabout, both human road users and automated vehicles must accomplish a series of tasks. First of all, they must perceive all other road users around them such as vehicles or vulnerable road users (VRUs). Based on perceived movements, intentions are estimated, e.g. which exit a vehicle is going to take. Finally, a trajectory is planed for entering the roundabout according to predicted gaps. Similar tasks have to be mastered while passing through and leaving a roundabout. At roundabouts with high traffic density or multiple lanes the degree of interaction is even higher. In summary, roundabouts are a very demanding driving situation that challenges many components of an automated vehicle at the same time. Thus, naturalistic recordings of this highly interactive behavior are valuable for development and testing of automated vehicles.

In this publication we present the rounD (**roun**dabout **d**rone) dataset, for which we used similar drone-based methods for recording and track extraction as for the highD [6] and inD [7] datasets (see Fig. 2). The dataset consists of the trajectories of road users that were extracted from traffic recordings at three different roundabouts in Germany. A cutout of a single frame of the dataset is shown in Fig. 1. As there is a lot of interaction such as negotiations between road users in the proximity of roundabouts, the dataset shall serve as a basis for the research of road user interaction, prediction and driver models.

II. Related Work

In recent years, several datasets for solving tasks in the domain of automated driving have been published. These can be roughly classified into perception and trajectory datasets. However, trajectories of road users are not contained in perception datasets like KITTI [8] or Waymo Open Dataset [9] and can, if at all, only be extracted from the perception data in a very limited and complex way. Therefore, we focus in the following on trajectory datasets.

For the collection and creation of trajectory datasets, several approaches have been pursued in recent years, namely the use of measurement vehicles [4], [10], infrastructure sensors [5] and measurements from an aerial perspective e.g. using camera-equipped drones [6], [7]. As shown in [6], a drone has many advantages such as high time and cost efficiency, minimal problems caused by perspective occlusion and negligible influence on traffic.

Drones and infrastructure sensors have particularly been used to generate datasets of road users on motorways as in [5], [6], [11] or in urban areas as in [7], [11], [12]. Hereafter, we focus on datasets containing at least one roundabout, which is the central topic of this publication.

The **Stanford Drone Dataset** [13] was published in 2016 and is one of the first datasets of road user trajectories including roundabouts. At in total eight recording sites on the Stanford university campus area, road users were recorded using a drone and extracted by a tracking algorithm. Two of the locations include roundabouts at which mostly bicyclists (1748) and pedestrians (1036) are present. Vehicles (260) are only present in the dataset to a very small extent. Furthermore, as can be seen in the videos provided, the resolution of the recorded videos is quite low and detections imprecise, resulting in noisy extracted trajectories. Finally, as all recording sites are located on a private campus, the recorded behavior is probably different to that on public roads.

The **Five Roundabouts Dataset** [4] is a dataset published in 2019. Using in total six Ibeo lidar scanners onboard of a vehicle that was parked near the roundabouts, more than 23 000 vehicles were tracked at five unsignalized roundabouts in Australia. As the focus of their work is the analysis of behavior of vehicles at roundabouts, no pedestrians were tracked for this dataset. Since the tracking system was suffering from a limited detection range and

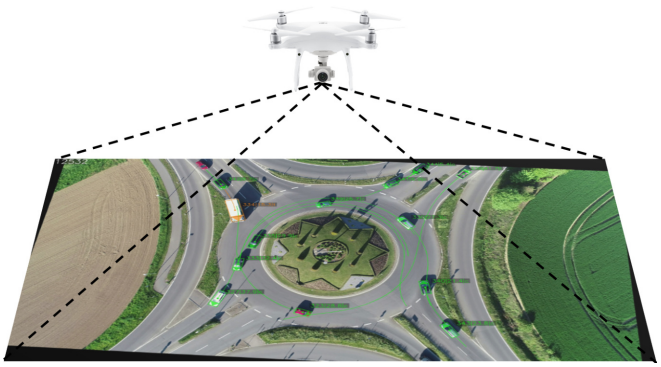


Fig. 2: We propose to use a camera-equipped drone to record traffic at roundabouts. The trajectory and class of each road user is extracted using computer vision algorithms.

perspective-related occlusions, vehicles entering from side roads are not considered. Finally, the quality of the created tracks varies, as according to the authors the lidar system had problems estimating the objects' size in case of complex vehicle shapes.

In the recently published **Interaction Dataset** [11], either a drone or static infrastructure sensors were utilized to track road users in different scenarios. A total of five different roundabouts in three countries (Germany, USA, China) were considered, at which 10 450 vehicles were tracked at a sampling frequency of 10 Hz. However, no statement is given about the accuracy of the provided data.

III. Methodology

Our method consists of a two-step process: Firstly, recording the naturalistic behavior of road users using a drone. Secondly, processing the recordings on a computing cluster, in which road users are detected in every video frame and tracks are created from these detections.

A. Recordings

All recordings were made under good weather conditions with adequate lighting and little wind to optimize the overall recording quality. This increased image sharpness and steadiness, facilitating further processing. A DJI Phantom 4 Pro was used for all recordings, which is equipped with a camera with a 4K resolution (4096x2160 pixel). The recordings were taken at maximum bitrate with 25 frames per second. During the recordings, the drone was hovering at a fixed flight altitude of 100 meters, which made it possible to cover the traffic on the roundabout itself and on parts of the roads connected to it. The size of a single pixel on the road surface is about 4x4 centimeters and the total covered area is approximately 140x70 meters. Depending on the measurement location and the wind conditions, each recording has a typical length between 19 and 23 minutes.

However, the recordings could not be used directly for processing without further pre-processing. Firstly, the recorded images are affected by typical camera lens distortions. To correct these, the drone's camera was calibrated and the resulting calibration parameters were applied to the video recordings. Secondly, even light wind and/or GPS drift during the recordings caused the drone to move or rotate slightly. However, to ensure that the static scene has a constant position in every frame, the recordings were stabilized on the first frame. To do so, a transformation estimator is used to find a projection from every frame to the first frame.

B. Detection

For the creation of tracks, we use a tracking-by-detection approach as in [7]. This means that every road user must be annotated with a polygon and the corresponding class in the frames in which they appear. As this is not manually doable, we utilized a neural-network-based detection approach. Simpler recognition-based approaches like SSD [14] or YOLOv3 [15] only generate axis-oriented bounding boxes. Because these do not tightly envelope the road users and do not include their rotation, they are not suitable. We have found that also more complex approaches like MaskRCNN [16] create semantic segmentations that are too coarse to derive accurate tracks.

Hence, we used a state-of-the-art semantic segmentation network architecture called DeepLab-v3+ [17], which is able to precisely assign every pixel of the 4K frames to a semantic class. From these semantic segmentations, detections can automatically be derived by thresholding the segmentations, grouping pixels based on the assigned classes and creating a polygon envelope for each detected road user. For training the network, we manually annotated road users in randomly selected frames by a polygon and one of the classes pedestrian, bicyclist, motorcyclist, car, van, truck, bus or trailer.

C. Tracking and Post-Processing

After the road users were detected in each frame individually, the detections needed to be matched by position and size in between consecutive frames to create tracks. As some of the road users like a pedestrian behind a street sign pole were temporarily not visible, the corresponding tracks were reassigned automatically by a Kalman-filter-based prediction or manually in a refining step.

After the tracking was done for all road users in a given recording, the tracks were post-processed to achieve smooth results. An RTS smoother [18] derived the position, direction, speed and acceleration for each timestamp of a track, taking into account the previous and subsequent time steps. Moreover, this also allowed to interpolate the trajectory of temporarily not visible road users.

In a last processing step, all tracks were converted from an image-based coordinate system to a local UTM-like coordinate system with a constant origin for all recordings at the same location. Thus, all recordings from the same recording location can simply be overlaid for analysis.

Finally, all tracks were drawn onto the recordings and checked manually. Errors like the assignment of a wrong class were corrected.

D. Dataset Format

For providing the dataset, we use a dataset format similar to the one used in [7]. We created both, files describing the recording site as well as files describing the recordings and extracted tracks. In order to describe a recording site, an image of the location, and a lanelet2 map [19] including driving lanes, traffic signs and traffic rules are given in the local UTM coordinate system. These information are necessary to consider the dependency of a road user's behavior on the static surroundings for analysis or e.g. for movement prediction.

For the tracks itself, in total three CSV files are provided for each recording describing the data on a different granularity. The first file gives information about the recording itself, including the location id, weekday, time of day or the total number of extracted road users among other information. The second file lists all road users in a recording including e.g. their class, size or when they are visible in the image frame.

Fig. 3: Map of recording sites in and around Aachen, Germany

Finally, the last file consists of the detailed trajectories for each road user. These include the position, heading, speed and acceleration for every time-step in the x and y direction of the static UTM coordinate systems as well as in the longitudinal and lateral movement direction of each road user.

Python scripts for handling the provided files and creating a visualization are provided at https://www.github.com/ika-rwth-aachen/drone-dataset-tools.

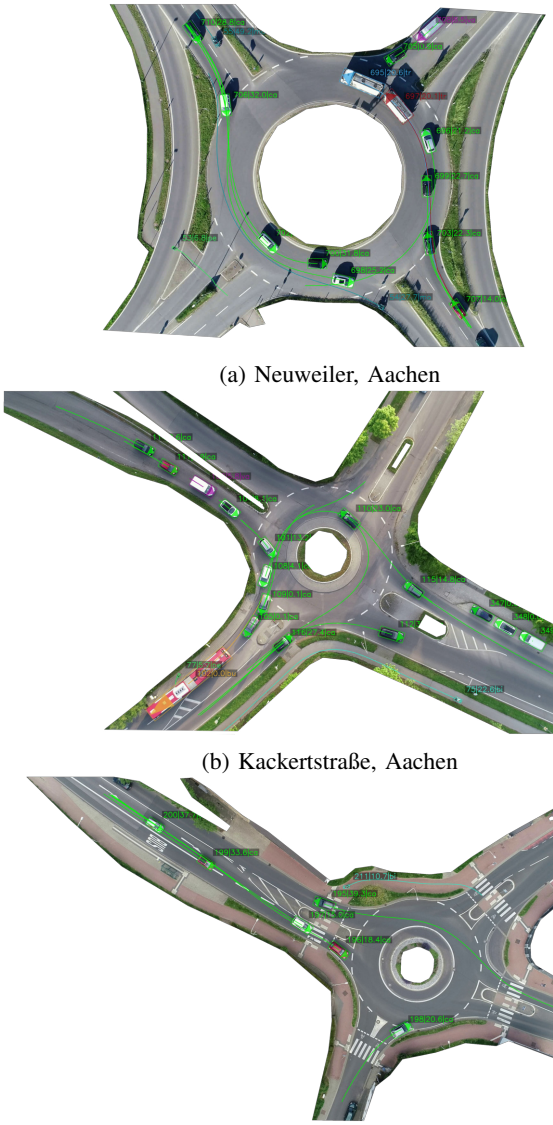
IV. RounD Dataset

A. *rounD at a Glance*

With the help of the pipeline described above, we have created the rounD dataset. For this, we took 24 separate recordings (more than 6 hours of video) of the traffic at three different measurement locations. Most of the recordings were made in the mornings to capture a high number of road users and a lot of interaction. From these recordings we were able to extract a total of 13 746 road users, which are more than 99% of the road users visible in the recordings. The most numerous recorded classes are cars (11 530), trucks (1061) and vans (608). Less frequent classes are trailers (257) and buses (53). VRUs including Pedestrians (25), bicyclists (88) and motorcycles (124) are less frequent due to the fact that the roundabouts are not located close to city centres or shopping areas. Despite the high degree of interaction at roundabouts, no collisions were recorded.

The dataset contains in total three recording sites located in and around Aachen in the western part of Germany as shown in Fig. 3, which can be characterized as follows:

Neuweiler is a four-armed roundabout that connects a highway with Aachen. This is the site with the highest traffic volume and therefore the site where thus most recordings



(a) Neuweiler, Aachen

(b) Kackertstraße, Aachen

(c) Thiergarten, Alsdorf

Fig. 4: Sample images of the three recording sites included in the rounD dataset. In each image, detections are annotated by colored bounding boxes (green: vehicles, purple: vans, light blue: trucks, red: trailers, orange: buses, light-green: pedestrians, dark-blue: motorcyclists, cyan: bicyclists).

were done. Due to the high traffic volume, there are additional lanes outside the roundabout allowing to skip the roundabout (see Fig. 4 (a)), which were not in focus due to the lack of interaction. In the roundabout itself there is a lot of interaction. Especially since all access roads are two-lane, while the exits are only single-lane. There are no lane markings in the roundabout itself, so that road users can perceive it as either single-lane or multi-lane. This difference in perception leads to interaction and a variety of scenarios in the roundabout.

Kackertstrasse is a roundabout in the urban area of Aachen and located in an industrial area. It connects a busy ring road (see top left in Fig. 4 (b)) around the city

TABLE I: Comparison of existing road user trajectory datasets for roundabouts

Dataset	Country	Locations	#Tracks	Road User Types	Quality	Map	Geo-Ref.	Sample Freq.
Stanford Drone	USA	campus (2)	3156	pedestrian, bicycle, car, skateboard, cart, bus	-	no	no	25 Hz
Five Roundabouts	Australia	urban (5)	∼23 000	bike, car, truck	o	no	no	25 Hz
Interaction	USA, Germany, China	not stated (5)	10 450	car	?	yes	no	10 Hz
rounD	Germany	(sub-)urban (3)	13 746	car, van, truck, bus, trailer, pedestrian, bicycle, motorcycle	+	yes	yes	25 Hz

with smaller roads leading into the city. Thus, most of the traffic flows through the arm connected to the ring road. The individual access and exit roads are all single-lane. In addition, there are traffic islands at the three less frequented arms where pedestrians cross the road. At the bus stop, buses occasionally block the traffic as people get on and off.

Thiergarten is a four-arm roundabout in a suburb of Aachen. All entrances and exits of this less frequented roundabout are single-lane. In addition, there is a traffic island on each arm and a crosswalk for passing pedestrians as shown in Fig. 4 (c).

B. Comparison with Existing Datasets

Even though the weighting of strengths and weaknesses of datasets depends on the exact purpose for which they are used, in the following we try to present a comparison of the existing datasets with the rounD dataset as generally as possible. In doing so, we address the aspects of size, quality and additional information provided as shown in Tab. I. For each dataset, we limit the analysis to the roundabouts.

In terms of **size and diversity**, it can be seen that the Stanford Drone Dataset not only contains the fewest tracks but also the fewest locations that are moreover located on private property. The remaining datasets show a mixed picture. The Interaction dataset contains sites from the most countries, but the fewest tracks and only a single type of road users (cars). The Five Roundabouts dataset as well as rounD have a larger number of tracks, whereas rounD additionally contains the most types of road users.

For the evaluation of the **quality**, we consider the information in the publications and available video material regarding detection rates, positioning accuracy, classification quality and sampling rate. For the Stanford Drone Dataset, the available videos show that most road users are tracked, but the size of the bounding box is inaccurate and varying. In addition, the drone recordings are not robustly stabilized because the static environment in the videos visibly moves. For the Five Roundabouts Dataset the authors describe that the sensors sometimes had problems to detect vehicles that are located in the peripheral area or have a special shape e.g. of the rear. This is mostly due to the lidar-based detection system used. The positioning of the sensors on a vehicle also resulted in vehicle-vehicle occlusions. For the Interaction dataset, it can be seen that it provides the lowest sampling rate of the datasets and does not indicate any vehicle acceleration of the tracks in the data. However, the quality of the other aspects could not be evaluated more closely, since no videos are public and no statement about the accuracy is made. The rounD dataset is the only dataset that recognizes a large number of road user types and also contains trailers as separate tracks. This is especially important for roundabouts, as a single bounding box does not adequately describe a vehicle with a trailer in a curve. Furthermore, the quality of the rounD data is high compared to the Stanford Drone Dataset and the Five Roundabouts dataset, which is due to the high resolution camera, the advantageous perspective and state-of-the-art algorithms. Thus, not only the positions but also speeds and accelerations could be derived precisely for more than 99% of the recorded road users.

If one compares the **additional data** provided for each dataset, one finds that the Five Roundabouts dataset provides trajectories in a normalized coordinate system to abstract between different roundabouts but no detailed information about the recording sites themselves. The Stanford Drone Dataset, on the other hand, provides reference images of the individual recordings, from which further information can be extracted, such as the position and size of elements like stop lines or sidewalks. Only the Interaction and rounD dataset provide HD maps in the lanelet2 format. This is especially important to take the static environment into account, e.g. when analyzing or predicting the behavior of road users. The rounD dataset is the only one that provides the exact location of the recording sites and gives geo-referenced coordinates. This is essential for extending the data with additional information from other maps if needed. Finally, weekdays and recording times are given in the rounD dataset.

V. Conclusion

In this paper we have shown the need for a road user trajectory dataset at roundabouts based on their entailed high interactivity. We created a deep-learning-based pipeline for extracting road users including vehicles and VRUs from drone recordings at roundabouts. Using this pipeline, we created the rounD dataset from more than six hours of recordings at three different recording sites in Germany. With more than 13 700 precisely tracked road users divided in eight different classes, rounD is the largest public dataset of road user trajectory at roundabouts on this quality level.

REFERENCES

[1] A. Kuefler, J. Morton, T. Wheeler, and M. Kochenderfer, "Imitating driver behavior with generative adversarial networks," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, pp. 204–211.

[2] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, and S. Savarese, "Sophie: An attentive gan for predicting paths compliant to social and physical constraints," in *2019 IEEE/CVF Conf. Computer Vision Pattern Recognition (CVPR)*, pp. 1349–1358.

[3] J. Langner, J. Bach, L. Ries, S. Otten, M. Holzäpfel, and E. Sax, "Estimating the uniqueness of test scenarios derived from recorded real-world-driving-data using autoencoders," in *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1860–1866.

[4] A. Zyner, S. Worrall, and E. M. Nebot, "Acfr five roundabouts dataset: Naturalistic driving at unsignalized intersections," *IEEE Intell. Transp. Syst. Mag.*, vol. 11, no. 4, pp. 8–18, winter 2019.

[5] US Department of Transportation. (2008) Ngsim - next generation simulation. [Online]. Available: https://ops.fhwa.dot.gov/trafficanalysistools/ngsim.htm

[6] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems," in *2018 21st Int. Conf. Intelligent Transportation Systems (ITSC)*, pp. 2118–2125.

[7] J. Bock, R. Krajewski, T. Moers, L. Vater, S. Runde, and L. Eckstein, "The ind dataset: A drone dataset of naturalistic vehicle trajectories at german intersections," 2019, arXiv:1911.07602 [cs].

[8] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.

[9] P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, V. Vasudevan, W. Han, J. Ngiam, H. Zhao, A. Timofeev, S. Ettinger, M. Krivokon, A. Gao, A. Joshi, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov, "Scalability in perception for autonomous driving: Waymo open dataset," 2019, arXiv:1912.04838 [cs].

[10] M.-F. Chang, J. W. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, and J. Hays, "Argoverse: 3d tracking and forecasting with rich maps," in *Conf. Computer Vision Pattern Recognition (CVPR)*, 2019.

[11] W. Zhan, L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kummerle, H. Konigshof, C. Stiller, A. de La Fortelle, and M. Tomizuka, "Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps," 2019, arXiv:1910.03088 [cs].

[12] T. Kucner, J. Saarinen, M. Magnusson, and A. J. Lilienthal, "Conditional transition maps: Learning motion patterns in dynamic environments," in *2013 IEEE/RSJ Int. Conf. Intelligent Robots Systems (IROS)*, pp. 1196–1201.

[13] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory understanding in crowded scenes," in *14th European Conf. Computer Vision (ECCV)*, 2016, pp. 549–565.

[14] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *14th European Conf. computer vision (ECCV)*, 2016, pp. 21–37.

[15] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," 2018, arXiv:1804.02767 [cs].

[16] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *2017 IEEE Int. Conf. Computer Vision (ICCV)*, pp. 2961–2969.

[17] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *15th European Conf. Computer Vision (ECCV)*, 2018, pp. 801–818.

[18] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic systems," *AIAA J.*, vol. 3, no. 8, pp. 1445–1450, 1965.

[19] F. Poggenhans, J.-H. Pauls, J. Janosovits, S. Orf, M. Naumann, F. Kuhnt, and M. Mayr, "Lanelet2: A high-definition map framework for the future of automated driving," in *2018 21st Int. Conf. Intelligent Transportation Systems (ITSC)*, pp. 1672–1679.