RoboBus: A Diverse and Cross-Border Public Transport Dataset

Georgios Varisteas, Raphaël Frank and François Robinet
Interdisciplinary Centre for Security, Reliability and Trust (SnT)
University of Luxembourg, 29 Avenue J.F Kennedy, L-1855 Luxembourg
firstname.lastname@uni.lu

***Abstract*—Academic datasets are an important source of information to validate and benchmark novel research concepts. In this paper we present RoboBus, a dataset recorded with a commercial bus on a cross-border public transport route between Luxembourg and France. The dataset contains approximately 8 hours of driving data divided into 15 trips that have been recorded over 4 days. It includes about 1.7 million anonymized images captured by two road-facing cameras, GNSS traces, data from a 9-axis IMU, and information directly retrieved from the CAN interface of the vehicle including speed, steering angle and position of the accelerator/brake pedals. We use an end-to-end autonomous driving approach that relies on imitation learning as use case example for the dataset.**

***Index Terms*—Dataset, Bus, Cross-border, Imitation Learning**

I. Introduction

The research on *Connected and Automated Driving* (CAD) requires vast amounts of real-world data to test, validate and benchmark new machine learning approaches. Over the last couple of years, multiple datasets have been published for different use-cases. The majority have been recorded with instrumented vehicles, often driven by the researchers themselves [1], [2], [3]. To the best of our knowledge their exists no dataset that has been recorded by a commercial bus operating on a cross-border public transport route. The RoboBus dataset presented in this work addresses this niche area. The recording system was designed to be integrated into a commercial bus without hindering its operation. This integration allowed us to collect data directly from the vehicle via the CAN interface, while it was being operated in a predefined location. This data includes speed, steering wheel angle and the position of the accelerator and brake pedals. Further, two cameras were mounted, one on the left and one on the right of the windshield. Finally, data from a GNSS and a 9-axis IMU were also recorded.

The dataset was collected on a cross-border route between Luxembourg and France over a period of 4 days. It is diverse in the sense that it captures different day times, weather conditions, traffic situations, road types and signalisation (different from one country to another).

We show that the collected data can be used to accurately train an end-to-end model to predict steering angles based on a single input image. This imitation learning method was first introduced by Pomerleau [4] and can be used as one building block of a self-driving vehicle.

The total dataset size has a size of 67GB and is available for download on GitHub [5].

The reminder of the paper is divided as follows. In Section II we introduce the related work. The RoboBus platform is introduced in Section III, followed by the dataset description in Section IV. A use-case example is presented in Section V and a conclusion is drawn in Section VI.

II. Related Works

Over the last decade, a number of datasets have been made available to the research community. The well known KITTI dataset [1] is a popular choice to benchmark computer vision tasks such as optical flow, visual odometry / SLAM and object detection for autonomous driving application. It has been recorded in Karlsruhe Germany using as primary sensor suite two cameras for stereo vision. The ground truth is provided by a LiDAR and GPS system. Over the years the dataset has been extended and new benchmarks added [6], [7]. Similarly the Cityscapes dataset [8], including street scenes from fifty different cities and containing a diverse set of annotated frames, is intended to asses the performance of vision algorithms on a defined set of tasks, and provide ground truth data to train deep neural networks. Those datasets are relatively small compared to the BDD100K dataset [3]. Recorded in different locations in the United States, it aims to provide large-scale, diverse road data with temporal information. The resulting dataset is composed of 120 million images and a rich set of annotations, including road objects, drivable areas, lane markings and full-frame instance segmentation.

When it comes to 3D object annotations, LiDAR data is usually included as ground truth. The nuScenes dataset [9] includes data from a diverse sensor suite, including on LiDAR, radars and cameras. It is diverse as it has been recorded in Singapore and Boston, both cities having different traffic rules. Likewise the large-scale and diverse camera-LiDAR dataset published by Waymo [10] contains annotated data with 2D and 3D bounding boxes.

Another approach was used to create the Oxford RobotCar dataset [2]. Here, no annotations are provided instead the same $10km$ route is driven twice a week for over a year with a vehicle equipped with multiple cameras, LiDARs and a GNSS. The aim was to capture changing traffic and weather conditions over a longer time period. A number of other datasets to

benchmark different tasks can be found on the Open Datasets website [11].

The previously mentioned datasets have all as primarily purpose to benchmark computer vision tasks. Another interesting concept is to use imitation learning to teach a vehicle to drive autonomously on a given route. To train such models a large amount of human driving data is needed. This includes data retrieved from the CAN interface of the vehicle, such a steering wheel angle and the position of the acceleration and braking pedals. There are only a limited number of such datasets available. The comma2k19 dataset [12] is one attempt in that direction. It contains 33 hours of commute on a highway in California and includes camera and CAN data.

The dataset presented in this paper has the following novelties: (1) it is, to the best of our knowledge, the first diverse dataset retrieved from a bus operating on a public transport route; (2) it is cross-border and captures the differences in road infrastructure and signage between two countries, (3) it contains lateral, (i.e. Steering Angle) and longitudinal (i.e. Acceleration/Brake Pedal position) control information retrieved directly from the CAN interface of the bus.

III. The RoboBus Platform

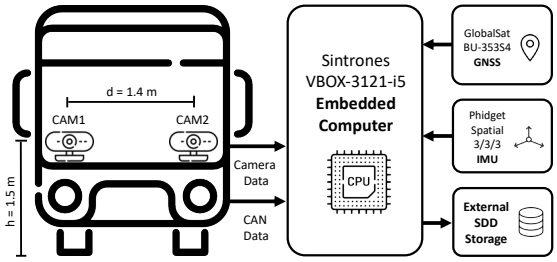

Fig. 1: RoboBus System Architecture.

The central element of the RoboBus platform is the Sintrones VBOX-3121-i5 embedded computer. It is connected to the CAN interface of a SETRA S431DT Ligne Double Decker bus. The *Fleet Management System Version 4* (FMS4)[1] protocol is used to retrieve vehicle data via this interface. The system connects to two Logitech C920 HD Webcams mounted on the left and right of the windshield at a height of approximately $1.5m$ from the ground, and a relative distance of $1.4m$. Further, a GlobalSat BU-353S5 GNSS receiver and a Phidget Spatial 3/3/3 (9-axis) IMU are connected to the embedded system. Finally an 1TB external Solid-State Drive (SSD) is used to store the collected data. The system is powered on and off via the ignition of the vehicle. The architecture of the platform is depicted in Figure 1.

IV. The RoboBus Dataset

A. Data Retrieval

The dataset was retrieved in June and July 2020 on a public transport route between Luxembourg City and Thionville in France[2]. Only the cross-border area depicted in Figure 2 has been considered for the data collection. The total distance of a one-way trip is $27km$, $12km$ on the Luxembourg side and $15km$ in France, including 10 bus stops. The route covers different road types including, highway, urban and country roads. Further, the $3km$ highway stretch contains a $600m$ long tunnel, where no GNSS information is available. The recording was performed on different week days covering different day times, traffic and weather conditions. The dataset contains a total of 15 distinct trips recorded of 4 days, 8 of which are southbound and 7 northbound.

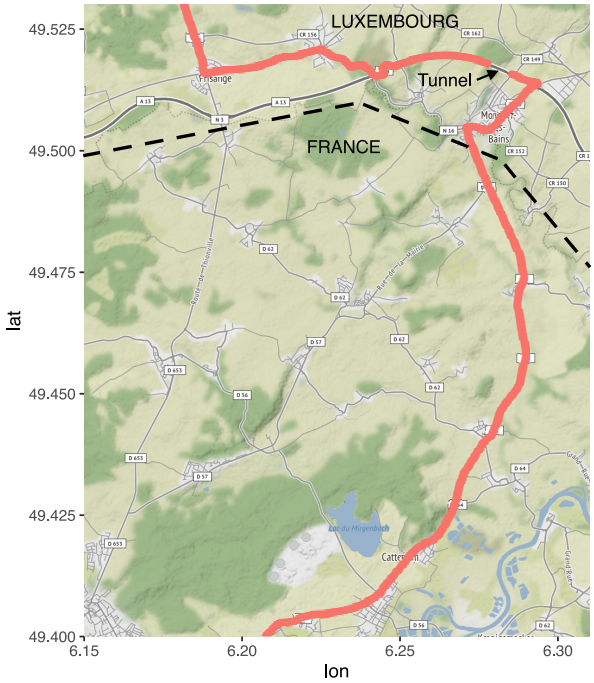

Fig. 2: Monitored Cross-border Area between Luxembourg and France (Bounding Box Coordinates: 6.15, 49.40, 6.31, 49.53).

B. Data Properties and Format

The initial release of the dataset has a total size of 67GB and is available for download on GitHub [5]. It contains roughly 1.7 million images recorded by two road-facing cameras with a frame-rate of $30fps$, which amounts to a total driving time of approximately 8 hours. Please note that during post-processing all images have been anonymized by blurring faces and license plates using the software made available by *understand.ai* [13]. Figure 4 illustrates images of diverse driving situations found in the dataset.

Table I summarizes the properties of the data as found in the CSV files for each trip with each line having the format: `[ID] [TIME_IN_MS] [DATA]`. The data received from the CAN interface has a variable frequency. The RoboBus recorder software has been configured to listen for the specified CAN

[1]https://bit.ly/3fZzfZ7

[2]https://bit.ly/2JonGPm

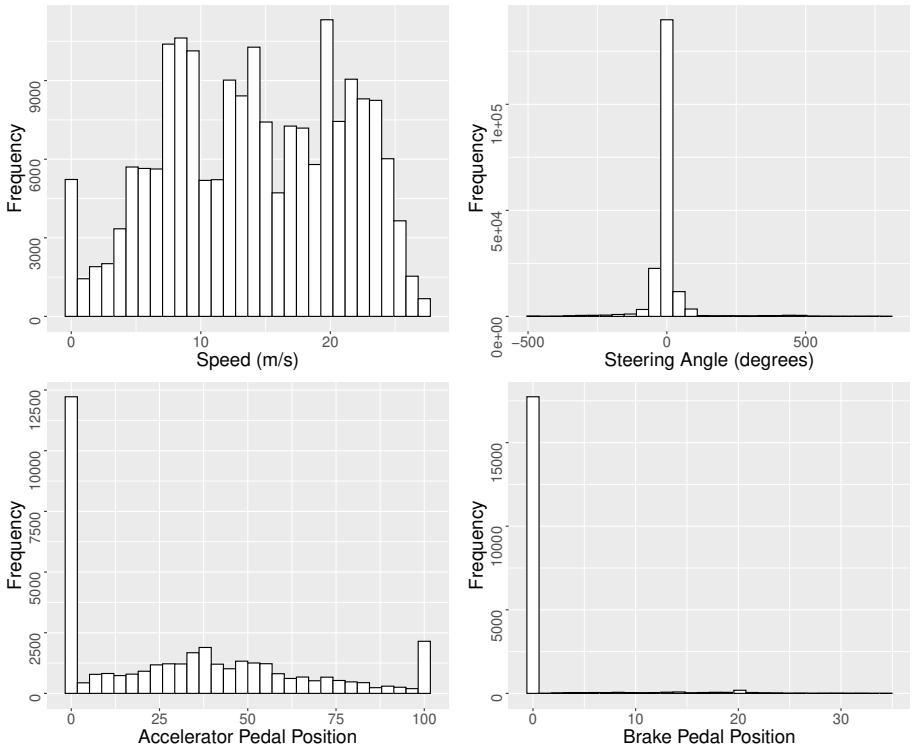

Fig. 3: Distributions of speed, steering wheel angle and accelerator, brake pedal position for a single trip.

Fig. 4: Sample images found in the dataset.

messages every millisecond. The rate at which the data is received depends on the bus type and the amount of traffic on the CAN interface.

Figure 3 shows the histograms of the speed, steering angle and acceleration/brake pedal position distributions for a single trip retrieved from the CAN interface. As expected the speed and the accelerator pedal position have wider distributions, whilst the steering angle and brake pedal position show only little variance.

V. Use Case: Imitation Learning

A. *Learning Steering Angles from a Single Camera*

Imitation learning has successfully been applied to learn complex control skills directly from sensor data. One applica-

Sensor	Freq.	ID	Data Format
CAN	var.	1	Vehicle speed (km/h)
		2	Accelerator pedal position
		3	Brake pedal position
		4	Steering wheel angle α
IMU	8Hz	50	Accelerometer (x,y,z)
		51	Magnetometer (x,y,z)
		52	Gyroscope (x,y,z)
GNSS	1Hz	100	Timestamp Lat. Lon. Alt.
Cameras	30fps	200	CAM1: RGB JPEG 640x480
		202	CAM2: RGB JPEG 640x480

TABLE I: Data properties and format.

tion is end-to-end steering, where steering angles are learned from road-facing camera frames recorded during human expert demonstrations. Such end-to-end control systems based on neural networks have been under study for decades. The first attempt, dubbed ALVINN for *Autonomous Land Vehicle In a Neural Network* was published in 1989 and relied on a 2-layer fully-connected network to predict 45 possible steering directions from a 30x32 input image [4]. Given its relative computational simplicity and the small size of the synthetic training set used at the time, this system was surprisingly effective and managed to accurately steer the vehicle along the 400-meter test path.

Following breakthroughs in computational capabilities and network architectures, researchers from NVIDIA have revisited the problem with impressive success in 2016 [14]. Their approach differs from ALVINN in that they use a much larger model, and treat the problem as the regression of a single steering angle, rather than a classification into one of 45 possible directions. Their model, dubbed *PilotNet*, is a 9-layer deep Convolutional Neural Network (CNN) that is trained using single images recorded from 3 cameras placed at different angles from the road. The PilotNet system empirically demonstrates that end-to-end systems are able to learn the task of lane and road following without manual decomposition into human crafted features such as roads or lanes.

Many improvements to end-to-end steering have since been explored in the literature. In an attempt to exploit temporal relationships between successive frames and their corresponding steering angles, *Xu et Al.* have proposed a novel architecture combining Long Short-Term Memory (LSTM) and CNNs [15]. Another notable attempt at addressing the limitations of steering models trained from human demonstration is Conditional Imitation Learning, where the model is trained to respond to both a input frame and a command describing the current maneuver [16]. At test time, this allows to easily manage the high-level actions taken by the model, while letting the learned policy handle the fine control details.

Since our aim is to demonstrate one of the many possible use-cases of the RoboBus dataset, we will experiment with the standard imitation learning framework, using a single camera and the PilotNet architecture. The details this architecture are illustrated on Figure 5.

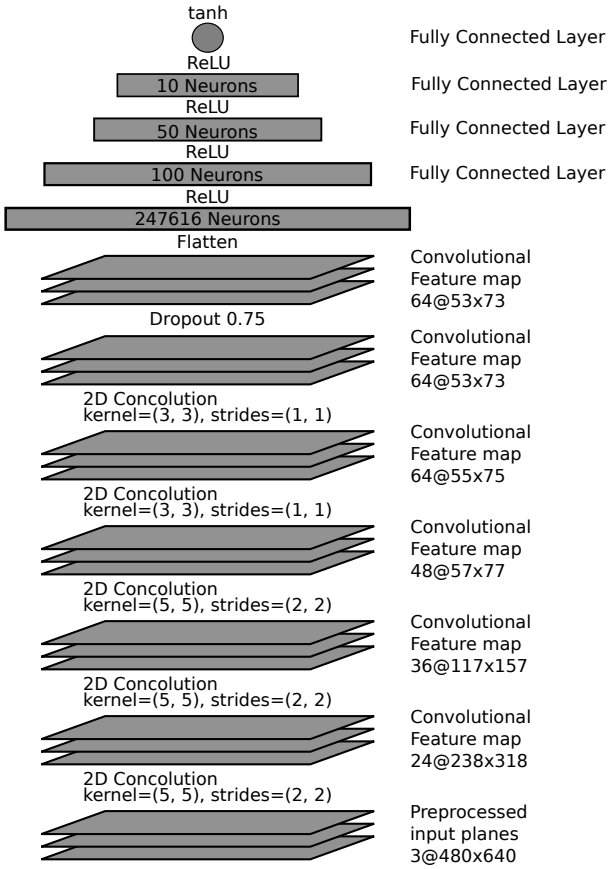

Fig. 5: Overview of the PilotNet architecture proposed by *Bojarski et Al*. Figure reproduced from [14].

Direction	Start time	Label
North	02:47:42	North 2
North	05:18:03	North 5
North	11:47:28	North 11
South	10:42:01	South 10
South	13:40:09	South 13

TABLE II: Trips used in the imitation learning use case. All subset of `RoboBus_11-6-2020_1.tar.gz`

B. Results

Our experiment uses 3 trips as training data, namely North 2, North 11, and South 13, and is evaluated on the separate North 5 trip, as explained in Table II. A random 20% split of the training set is reserved for validation. This results in 22400 training frames, and 5600 and 11700 validation and testing frames respectively.

Input frames x of resolution $480x640$ are converted to the YUV color space and their intensities normalized to the $[-1, 1]$ interval. To facilitate optimization and make the model less specific to our vehicle, ground truth steering angles α are also mapped from their original $[-650, 650]$ degree range to $[-1, 1]$. Correspondingly, a hyperbolic tangent activation is applied to the output of PilotNet.

The model parameters θ are optimized to to minimize the

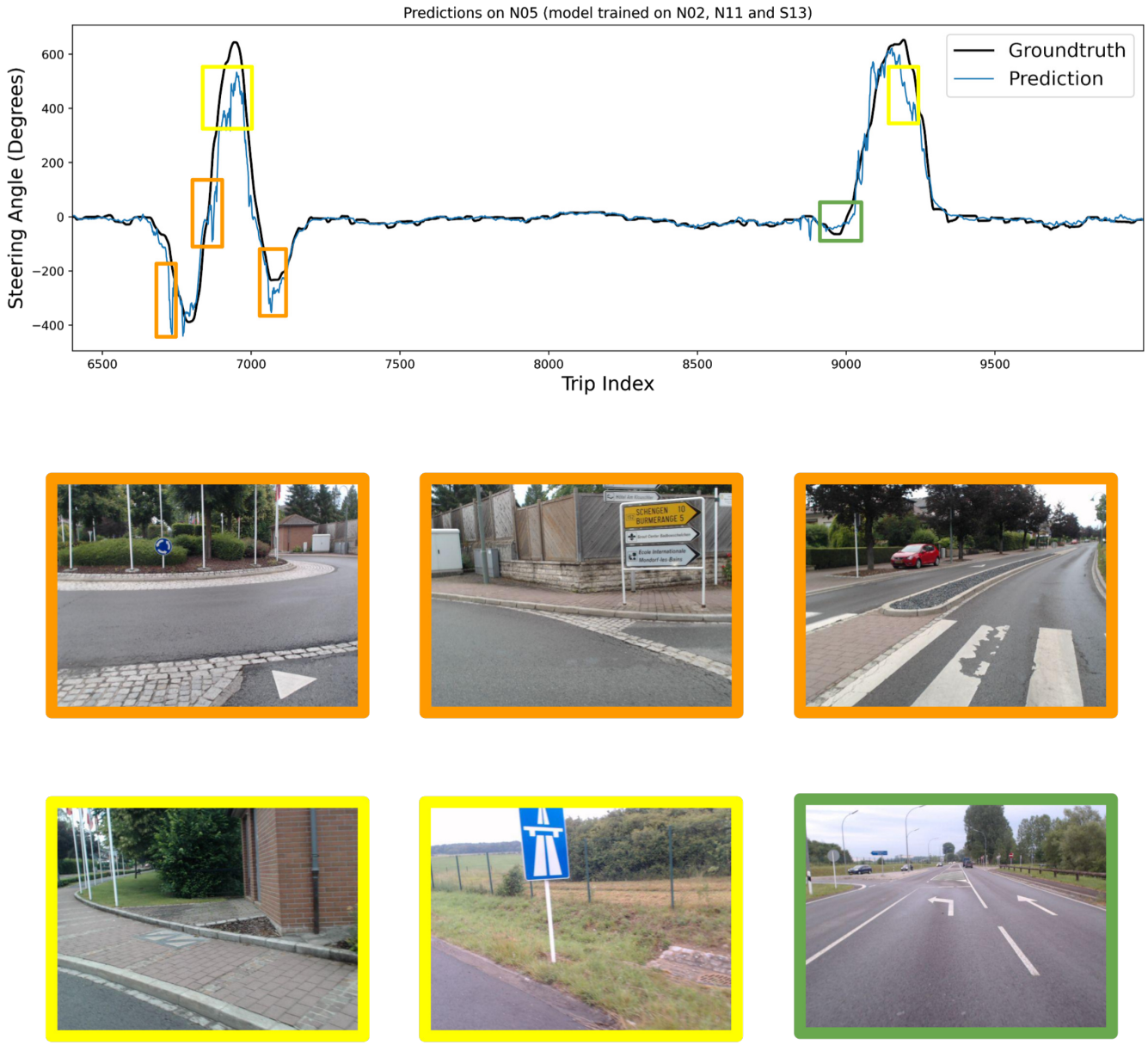

Fig. 6: Steering angles learned using a PilotNet model evaluated on sequence North 5. For each border color, the sample frames are shown in the order in which they appear in the sequence.

squared prediction error

$$\min_{\theta} \|f(x;\theta) - \alpha\|_2^2 .$$

The Adam optimizer is used with an initial learning rate of 10^{-3} and a batch size of 128 [17]. In an effort to reduce overfitting, features from the last convolutional layer are randomly dropped out with a probability of 0.3 while learning and training is stopped early, once the validation loss has stopped decreasing [18].

Figure 6 shows the predictions of our trained model on the test trip North 5. The mean squared test error of 0.006 could likely be further improved through additional tuning of hyper-parameters, architecture changes, or dataset specific preprocessing like area-of-interest cropping. However, obtaining a state-of-the-art steering model is not the focus of this work. Instead, we will focus our effort on analyzing the behavior of the learned model in several scenarios, outlined in different colors on Figure 6.

The orange frames are examples of cases where our model steers too much to the right. In the first and third images, specific road characteristics such as road pavement and a pedestrian crossing are likely throwing the model off. This could be prevented by leveraging a visualization of the network focus at training time to encouraging it to focus on the road itself [19]. In the second image, the model attempts to take a right turn while the human demonstrator only decided to leave the roundabout at the next exit. This mismatch can be fixed through conditional imitation learning, by including the intent of the demonstrator as an additional learning signal [16]. The yellow-bordered frames have been selected to illustrate under-steering scenarios in sharp turns. Although our model fails to reach the maximum steering values, it still predicts large angles in the correct direction in these situations. Finally, the green frame shows an example where PilotNet arguably surpasses the human demonstrator. When approaching the left turn, the human demonstrator first veers to the right before taking a sharper left turn, while our model approaches the same turn by smoothly veering to the left.

VI. Conclusion and Future Work

In this paper we presented RoboBus, a diverse dataset recorded by a bus operating on a cross-border public transport route between Luxembourg and France. It contains a large amount of reprocessed images recorded by two road-facing cameras as well as IMU, GNSS and CAN data. By using as an example imitation learning, we show that this data can be used to accurately train an end-to-end model to predict steering angles using frames coming from a single road-facing camera. We believe that RoboBus is a good addition to the previously published datasets due to its unique characteristics.

Future work will consist in adding more trips to the dataset and adding contextual information such as weather and traffic conditions.

Acknowledgments

This work has been financially supported by the EU INTERREG GR Terminal project. The authors would also like to thank Voyages Emile Weber Luxembourg for their support and for granting us access to their bus.

References

[1] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in Conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[2] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 Year, 1000km: The Oxford RobotCar Dataset," The International Journal of Robotics Research (IJRR), vol. 36, no. 1, pp. 3–15, 2017. [Online]. Available: http://dx.doi.org/10.1177/0278364916679498

[3] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving video database with scalable annotation tooling," CoRR, vol. abs/1805.04687, 2018. [Online]. Available: http://arxiv.org/abs/1805.04687

[4] D. A. Pomerleau, "Advances in neural information processing systems 1," D. S. Touretzky, Ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1989, ch. ALVINN: An Autonomous Land Vehicle in a Neural Network, pp. 305–313.

[5] "Robobus dataset," https://github.com/raphaelfrank/robobus, accessed: 2020-12-01.

[6] M. Menze and A. Geiger, "Object scene flow for autonomous vehicles," in Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[7] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," CoRR, vol. abs/1406.2283, 2014. [Online]. Available: http://arxiv.org/abs/1406.2283

[8] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[9] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuscenes: A multimodal dataset for autonomous driving," arXiv preprint arXiv:1903.11027, 2019.

[10] 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, S. Zhao, S. Cheng, Y. Zhang, J. Shlens, Z. Chen, and D. Anguelov, "Scalability in perception for autonomous driving: Waymo open dataset," 2020.

[11] "Open datasets - scale," https://scale.com/open-datasets, accessed: 2020-11-23.

[12] H. Schafer, E. Santana, A. Haden, and R. Biasini, "A commute in data: The comma2k19 dataset," 2018.

[13] "understand.ai anonymizer," https://github.com/understand-ai/anonymizer, accessed: 2020-12-02.

[14] M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, "End to end learning for self-driving cars," CoRR, vol. abs/1604.07316, 2016. [Online]. Available: http://arxiv.org/abs/1604.07316

[15] H. Xu, Y. Gao, F. Yu, and T. Darrell, "End-to-end learning of driving models from large-scale video datasets," in 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 3530–3538.

[16] F. Codevilla, M. Müller, A. López, V. Koltun, and A. Dosovitskiy, "End-to-end driving via conditional imitation learning," in 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018, pp. 4693–4700.

[17] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," International Conference on Learning Representations, 12 2014.

[18] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," J. Mach. Learn. Res., vol. 15, no. 1, p. 1929–1958, Jan. 2014.

[19] F. Robinet, A. Demeules, R. Frank, G. Varisteas, and C. Hundt, "Leveraging privileged information to limit distraction in end-to-end lane following," in 2020 IEEE 17th Annual Consumer Communications Networking Conference (CCNC), 2020, pp. 1–6.