Proceedings of the 2022 IEEE International Conference on Multisensor Fusion and Integration
(MFI 2022), 20-22 September, Cranfield, UK
DOI:10.1109/MFI55806.2022.9913844

Predicting Autonomous Vehicle Navigation Parameters via Image and Image-and-Point Cloud Fusion-based End-to-End Methods

Semih Beycimen, Dmitry Ignatyev and Argyrios Zolotas

***Abstract*— This paper presents a study of end-to-end methods for predicting autonomous vehicle navigation parameters. Image-based and Image & Lidar points-based end-to-end models have been trained under Nvidia learning architectures as well as Densenet-169, Resnet-152 and Inception-v4. Various learning parameters for autonomous vehicle navigation, input models and pre-processing data algorithms i.e. image cropping, noise removing, semantic segmentation for image data have been investigated and tested. The best ones, from the rigorous investigation, are selected for the main framework of the study. Results reveal that the Nvidia architecture trained Image & Lidar points-based method offers the better results accuracy rate-wise for steering angle and speed.**

I. INTRODUCTION

It is well know that AI/ML End-to-End learning refers to methods where the model learns all the steps between the initial input phase and the final output result. In the learning process the different parameters are trained simultaneously [1]. Referring to the topic of ground vehicle traversability assessment, such kind of methods are incorporated in several papers. End-to-end methods research mainly cover the terrain understanding and control tasks such as the prediction of steering angle or speed. End-to-end methods comprise *deep learning* and *reinforcement learning*- based approaches. Here, we employ the former methods to predict steering angle and speed of the vehicle.

Deep learning approaches receive substantial interest from the wider research community. Few resources are listed in this paper, that are relevant to the study. Work in [2] demonstrated an end-to-end method navigation behaviour and imaging approach for mobile robots. The robot navigation commands: *left*, *right*, and *forward* as well as images from three cameras (a hiker was equipped with the relevant sensors traversing the forest trail in the study) were collected to create the dataset. All images were labelled corresponding to the terrain class as a part of supervised learning methods. To predict the navigation command, deep learning (DNN) method was used. Validation was performed in real environment using a aerial robot (quadrotor). This was an innovative way to obtain the dataset, using a human (hiker) subset, however this is not realistic for off-road ground vehicle traversability.

A recent paper by [3] proposed a road detection method using a camera and Lidar (for on-road vehicle application). The contribution of the study was the fusion of information form thwo types of sensors, i.e. Lidar and cameras. Point clouds from Lidar were converted to the different feature maps, and the neural network used as inputs the feature maps and camera images. The authors presented different fusion algorithms, i.e. *early*, *late*, and (their proposed) *cross fusion* to compare information fusion effects. According to the analysis of their results, the best accurate model with 96.03% was obtained using the cross-fusion approach with the KITTI dataset[4]. Clearly their model shown good and promising performance, albeit the study and setup favors on-road vehicle scenarios.

Regarding vehicle traversability (or vehicle going) safety, a study by [5] implemented a safety solution to avoid collisions (the authors referred to their approach as Simplex-Drive). The study related to on-road vehicles, and the validation was using a lane changing scenario in dense traffic. Researchers have also proposed regression algorithm to predict exact navigation parameters, and combined methods have been used in various studies for improving prediction. The authors in [6] studied twelve driver actions such as *left turn*, *straight* and *right turn* predicted from a Driver Behavior Classification (DBC) algorithm, and steering angle from a Steering Angle Regression (SAR) algorithm. The camera image, Lidar data and odometry data have been used in the SAR algorithm as inputs (while only the camera image was used in DBC). Gated Recurrent Fusion Unit (GRFU) learning algorithm, similar to LSTM, were implemented for improving prediction accuracy. The authors verified the approach using Open Racing Car Simulator (TORCS)[7] and Honda Driving datasets[8] that have been gathered in the simulation and also using real environment, respectively. The proposed method was shown to have improved the Mean Squared Error (MSE) and mean Average Precision (mAP) score.

As mentioned, the literature is rich regarding end-to-end methods and only a small/selected set has been discussed here. The interested reader is referred to further resources in the literature that study this topic, in particular deep learning, i.e. [9], [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23].

This paper presents a study of end-to-end methods for predicting autonomous vehicle navigation parameters. Image-based and Image & Lidar points- based end-to-end models have been trained under Nvidia learning architectures as well as Densenet-169, Resnet-152 and Inception-v4. Results reveal that the Nvidia architecture trained Image & Lidar points- based method offers the best results having an accuracy rate of 73% and 84% for steering angle and speed,

The authors are with the Centre for Autonomous and Cyber-physical Systems, Cranfield University, SATM, Bedford MK43 0AL, United Kingdom; {semih.beycimen, d.ignatyev, a.zolotas}@cranfield.ac.uk

© 2022 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

respectively. Our approach is presented in the methodology section, and further discussion is presented in the results section.

II. DATASET

Regarding the dataset used in this study, we refer to the DBNET raw dataset [24]. We perform pre-processing of the data such as cropping, noise removing, semantic segmentation, downsampling of Lidar data, to obtain the following dataset distribution: Total Dataset= 26178, Train use= 16680 data (63.71%), Evaluation use= 4358 data (16.64%), Test use= 5140 data (19.63%).

III. METHODOLOGY

The main model determination phase is particularly important in training algorithms. Figure 1 presents the process chart. It can be seen that various parameters in processing or pre-processing steps such as *batch-size*, *learning rate*, *image cropping*, *data augmentation*, *data noise removal*, arranging point cloud to predict steering angle and vehicle speed. Pre-processing was an important step to select the parameters for the study. Moreover, Convolutional Neural Networks receive close attention when it comes to prediction of parameters. Four learning methods, namely *Nvidia*, *Inception-v4*, *Densenet-169* and *Resnet-152* have been used to predict steering angle and vehicle speed here.

Tolerances for calculation of accuracy have been selected at 0.05 for steering angle and 0.1 for speed. That means, If the predicted steering angle or speed value is lower than the tolerance, the algorithm is acknowledge it as a correct label. Accuracy is the rate of correct label values in the total labels.

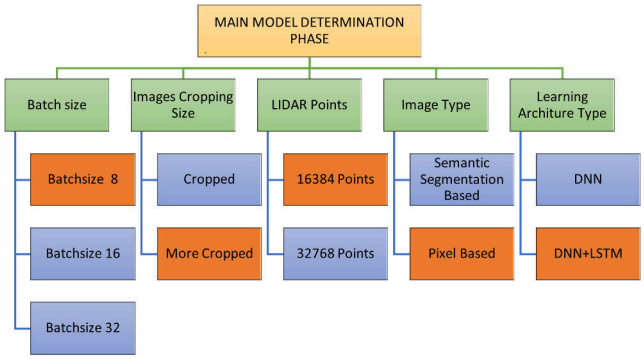

Fig. 1. Process Chart for Determining Main Parameters & Properties and Processing Algorithms - **Orange**: Selected after-test steps, **Blue**: Others

IV. MAIN MODEL DETERMINATION PHASE

In this section, we discuss the parameters for the main model determination phase.

A. Batch-size

Three batch-size models have been tested on the same model and parameters, the aim being to note the effect of batch-size on the learning algorithm. We noted that Batch-size= 8 is better in terms of the combination of: accuracy, training time, accuracy for test data and response time (accuracy graphs can be seen in Figure 2)

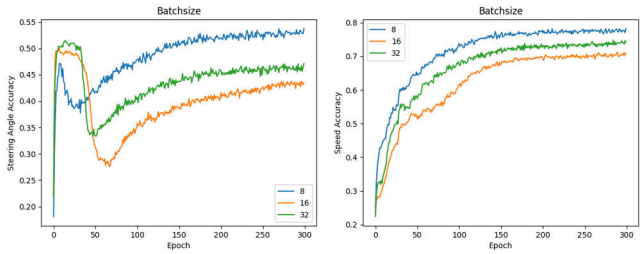

Fig. 2. Batch-size comparison (steering angle and speed accuracy)

B. Image Cropping

While vision sensors (cameras) provide much useful information for the surrounding environment of the vehicle, typically there are unnecessary parts of the image such as sky, or insignificant objects (per the application) which are normally removed to target training and improve results. We perform two cropping models, to obtain the optimised outcome, see Figure 4. Firstly, the images have been cropped 420 pixels left and right, and 240 pixels top and bottom (note that images were first converted to 840x600). For the second case, the images were resized to 760x400 an cropped. Once cropping completed the images were resized to 66x200, 224x224 and 299x299 for the learning algorithms. Figure 3 illustrates that the two models are almost identical (with minor differences). However, we select the second approach given the training time, computational power and slightly increased accuracy.

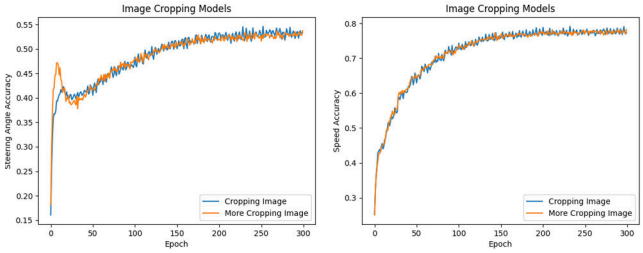

Fig. 3. Image cropping comparison (steering angle and speed accuracy)

C. Lidar Points Down-Sampling

Lidar sensor-wise there are almost 1 million points in the point cloud dataset (some portions between 300k-400k). Such large data set portion may hinder training of the learning algorithm due to the effect of large point cloud on time and computational power requirement. As a result, we opt to down-sampling the Lidar dataset to 16,384 and 32,768 points. We also apply SOR (Statistical Outlier Removal) Filter and Noise Filter (VoxelGrid) methods have been applied to remove noise from the point cloud (Figure 5. Moreover, Figure 6 compares the Lidar Points Down-Sampling model versions. IT is seen that accuracy is almost identical in the two cases, however training time has decreased to almost half using the 16,384 points model.

Fig. 4. Image Cropping model

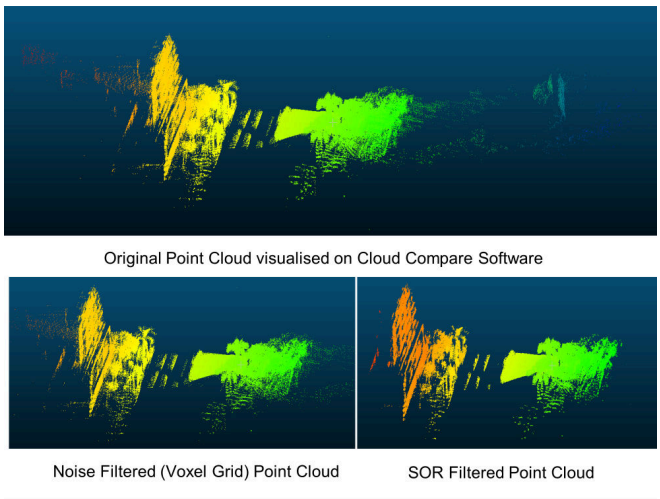

Fig. 5. Point Cloud Noise Filter Models

D. Image Type

We employ two image model types i.e. RGB image and Segmented image obtained from semantic segmentation algorithm (Figure 7) have been used in the training algorithms. We noted that segmented images effected the learning algorithm, typically links to complex environment and/or incorrect /inadequate labelling. Hence, the RGB image approach has been chosen for the main model of study.

E. Neural Network Type

We compare Deep Neural Network (DNN) and DNN+Long Short-Term Memory(LSTM) methods for the training of the algorithms. Adding LSTM to neural network has increased the accuracy rate more than 10% for the steering angle and 5% percent for the speed, refer to Figure 8.

V. RESULTS FOR MAIN ALGORITHMS

As discussed in the previous section, the best parameters for the main study were used to provide the results for the

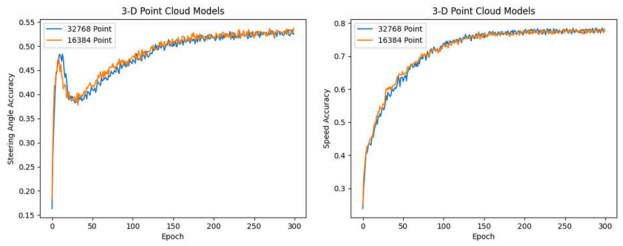

Fig. 6. Comparison of Lidar Points Down-Sampling Models

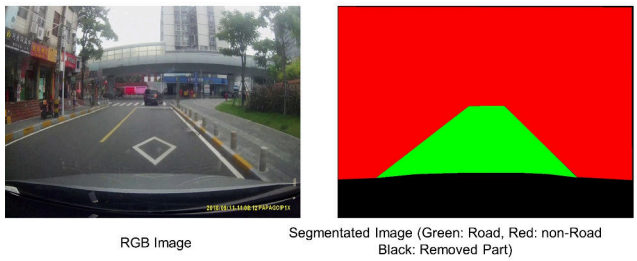

Fig. 7. Image Types Left: RGB Image Right: Segmented Image

main algorithms. The better cropped RGB image, the 16,384 Lidar points used as inputs and batch-size= 8, the learning rate was set to 0.001, DNN+LSTM network model have been used in the learning model. Hence, two end-to-end models (only image-based and image & Lidar point based) have been tested with four different learning models i.e. Inception-v4, Resnet-152, Nvidia and Densenet-169 (see Figure 9). For the training and results we used the Cranfield University HPC facility with V100 GPU card and two Intel E5-2620 v4 (Broadwell) CPUs.

Using the above setup, training was performed and tabulated to compare and contrast the methods and algorithms. As seen from Figure 8, Nvidia DNN+LSTM model provides the better result comprising the image & Lidar point-based input model. In this context, the accuracy rate achieved was an average of 73% and 84% for the steering angle and speed, respectively (this can also be noted on Table). Clearly incorporating the 3D Point Cloud to the model improved accuracy c.15%.

Also, actual values and predicted values from learned model for steering angle and vehicle speed have been compared under best model, Nvidia learning architecture with inputs Image & and Lidar Points. Model 1, that is only image

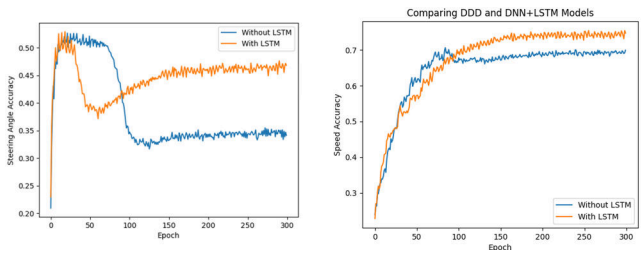

Fig. 8. Comparison of DNN and DNN+LSTM models

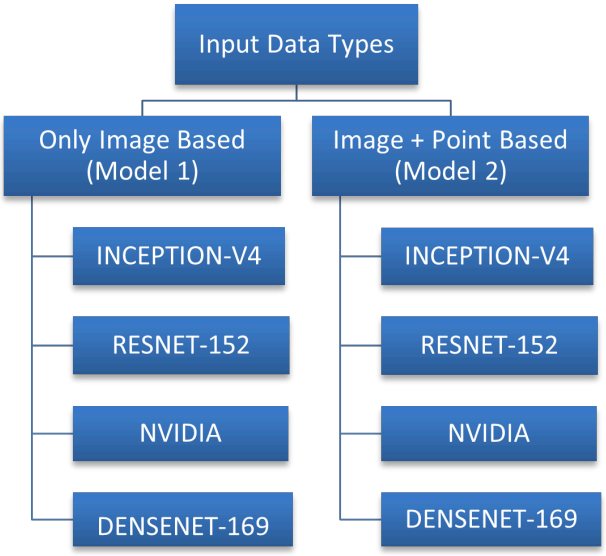

Fig. 9. Architecture of Main Approach

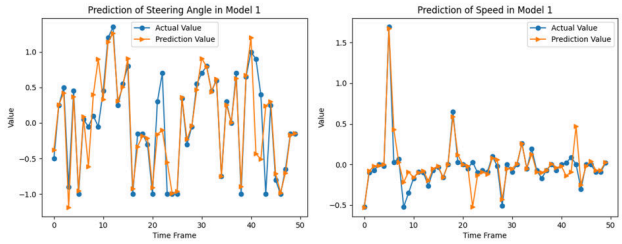

Fig. 10. Comparison of Actual and Predicted Values for Model 1

based input, has been demonstrated in Figure 10 and Model 2, fused data, in Figure 11. The Figures illustrate that Model 2 is more favorable.

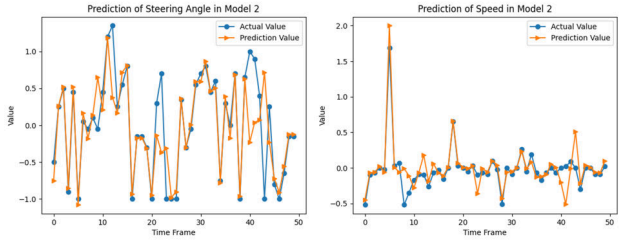

Fig. 11. Comparison of Actual and Predicted Values for Model 2

VI. CONCLUSIONS

This paper presented work which is part of a PhD study looking into advanced off-road ground vehicle traversability. In particular, the study in this paper refers to end-to-end methods for predicting autonomous vehicle navigation parameters. We presented rigorous investigation of Image-based and Image & Lidar points- (fused) based end-to-end models of Nvidia learning architectures, Densenet-169, Resnet-152 and Inception-v4. Various learning parameters

TABLE I
COMPARISON OF LEARNING ARCHITECTURES ACCURACY FOR MODELS 1 AND 2

Learning Architectures	Parameter Type	Model 1 Accuracy	Model 2 Accuracy
Inception-v4	Steering Angle	44%	61%
	Speed	68%	73%
Resnet-152	Steering Angle	50%	65%
	Speed	66%	78%
Nvidia	Steering Angle	68%	81%
	Speed	77%	88%
Densenet-169	Steering Angle	36%	55%
	Speed	54%	65%

for autonomous vehicle navigation, input models and pre-processing data algorithms i.e. image cropping, noise removing, semantic segmentation for image data have been investigated and tested. It was found that the Nvidia learning network provided more accurate and reliable results for predicting steering angle and speed for the vehicle. Adding an LSTM (a recurrent neural network (RNN) type) to the model improved accuracy considerably. Results of this paper, although with the on-road dataset, inform future work that relates to collecting further datasets using a UGV platform in off-road environments, further training the models and advancing the traversability solutions.

VII. CRediT AUTHORSHIP CONTRIBUTION STATEMENT

Conceptualisation: SB, DI, AZ. Methodology: SB, AZ. Writing – original draft: SB. Manuscript revision: DI, AZ. Supervision: DI, AZ.

VIII. DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

IX. ACKNOWLEDGEMENT

The first author acknowledges Republic of Turkey, Ministry of National Education (YLYS), for supporting the studies under PhD scholarship ref. U9BYTAB2LDGA7LK.

REFERENCES

[1] 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, K. Zieba, End to end learning for self-driving cars, CoRR abs/1604.07316 (2016). arXiv:1604.07316.
URL http://arxiv.org/abs/1604.07316

[2] A. Giusti, J. Guzzi, D. C. Ciresan, F. L. He, J. P. Rodriguez, F. Fontana, M. Faessler, C. Forster, J. Schmidhuber, G. D. Caro, D. Scaramuzza, L. M. Gambardella, A Machine Learning Approach to Visual Perception of Forest Trails for Mobile Robots, IEEE Robotics and Automation Letters 1 (2) (2016) 661–667. doi:10.1109/LRA.2015.2509024.

[3] L. Caltagirone, M. Bellone, L. Svensson, M. Wahde, LIDAR–camera fusion for road detection using fully convolutional neural networks, Robotics and Autonomous Systems 111 (2019) 125–131. arXiv:1809.07941, doi:10.1016/j.robot.2018.11.002.

[4] A. Geiger, P. Lenz, C. Stiller, R. Urtasun, Vision meets robotics: The KITTI dataset. The International Journal of Robotics Research, The International Journal of Robotics Research (October) (2013) 1–6.

[5] J. Grieser, M. Zhang, T. Warnecke, A. Rausch, Assuring the safety of end-to-end learning-based autonomous driving through runtime monitoring, in: 2020 23rd Euromicro Conference on Digital System Design (DSD), IEEE, 2020, pp. 476–483.

[6] A. Narayanan, A. Siravuru, B. Dariush, Gated Recurrent Fusion to Learn Driving Behavior from Temporal Multimodal Data, IEEE Robotics and Automation Letters 5 (2) (2020) 1287–1294. doi:10.1109/LRA.2020.2967738.

[7] B. Wymann, E. Espié, C. Guionneau, C. Dimitrakakis, R. Coulom, A. Sumner, Torcs, the open racing car simulator, Software available at http://torcs. sourceforge. net 4 (6) (2000) 2.

[8] V. Ramanishka, Y.-T. Chen, T. Misu, K. Saenko, Toward driving scene understanding: A dataset for learning driver behavior and causal reasoning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 7699–7707.

[9] Y. Shen, L. Zheng, M. Shu, W. Li, T. Goldstein, M. C. Lin, Improving Robustness of Learning-based Autonomous Steering Using Adversarial Images (2021). arXiv:2102.13262.
URL http://arxiv.org/abs/2102.13262

[10] A. Kumar, S. Palaniswamy, Steering angle estimation for self-driving car using deep learning, in: Symposium on Machine Learning and Metaheuristics Algorithms, and Applications, Springer, Singapore, 2019, pp. 196–207.

[11] C. J. Holder, T. P. Breckon, Learning to drive: End-to-end off-road path prediction, IEEE Intelligent Transportation Systems Magazine 13 (2) (2021) 217–221.

[12] S. Du, H. Guo, A. Simpson, Self-driving car steering angle prediction based on image recognition, arXiv preprint arXiv:1912.05440 (2019).

[13] V. Singhal, S. Gugale, R. Agarwal, P. Dhake, U. Kalshetti, Steering angle prediction in autonomous vehicles using deep learning, in: 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), IEEE, 2019, pp. 1–6.

[14] M. Islam, M. Chowdhury, H. Li, H. Hu, Vision-based navigation of autonomous vehicles in roadway environments with unexpected hazards, Transportation research record 2673 (12) (2019) 494–507.

[15] J. Jhung, I. Bae, J. Moon, T. Kim, J. Kim, S. Kim, End-to-end steering controller with cnn-based closed-loop feedback for autonomous vehicles, in: 2018 IEEE intelligent vehicles symposium (IV), IEEE, 2018, pp. 617–622.

[16] Z. Yang, Y. Zhang, J. Yu, J. Cai, J. Luo, End-to-end multi-modal multi-task vehicle control for self-driving cars with visual perceptions, in: 2018 24th International Conference on Pattern Recognition (ICPR), IEEE, 2018, pp. 2289–2294.

[17] S. C. Jugade, A. C. Victorino, V. B. Cherfaoui, S. Kanarachos, Sensor based prediction of human driving decisions using feed forward neural networks for intelligent vehicles, in: 2018 21st International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2018, pp. 691–696.

[18] J. Kim, J. Canny, Interpretable Learning for Self-Driving Cars by Visualizing Causal Attention, Proceedings of the IEEE International Conference on Computer Vision 2017-Octob (2017) 2961–2969. arXiv:1703.10631, doi:10.1109/ICCV.2017.320.

[19] L. Chi, Y. Mu, Learning end-to-end autonomous steering model from spatial and temporal visual cues, VSCC 2017 - Proceedings of the Workshop on Visual Analysis in Smart and Connected Communities, co-located with MM 2017 (2017) 9–16arXiv:arXiv:1708.03798v1, doi:10.1145/3132734.3132737.

[20] C. Sevastopoulos, K. M. Oikonomou, S. Konstantopoulos, Improving traversability estimation through autonomous robot experimentation, in: International Conference on Computer Vision Systems, Springer, 2019, pp. 175–184.

[21] J. J. Meyer, End-to-end learning of steering wheel angles for autonomous driving, Ph.D. thesis, Bachelor's Thesis, Freie Universität Berlin, Berlin, Germany, 2019. (2019).

[22] P. Kicki, T. Gawron, K. Ćwian, M. Ozay, P. Skrzypczyński, Learning from experience for rapid generation of local car maneuvers, Engineering Applications of Artificial Intelligence 105 (2021) 104399. doi:https://doi.org/10.1016/j.engappai.2021.104399.
URL https://www.sciencedirect.com/science/article/pii/S0952197621002475

[23] Álvaro Javier Prado, M. Michałek, F. Cheein, Machine-learning based approaches for self-tuning trajectory tracking controllers under terrain changes in repetitive tasks, Engineering Applications of Artificial Intelligence 67 (2018) 63–80. doi:https://doi.org/10.1016/j.engappai.2017.09.013.
URL https://www.sciencedirect.com/science/article/pii/S0952197617302270

[24] Y. Chen, J. Wang, J. Li, C. Lu, Z. Luo, H. Xue, C. Wang, Lidar-video driving dataset: Learning driving policies effectively, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 5870–5878. doi:10.1109/CVPR.2018.00615.