Imitative Models for Passenger-Scale Autonomous Off-Road Driving

Nitish Dashora[1*], Sunggoo Jung[2*], Dhruv Shah[1], Valentin Ibars[2], Osher Lerner[1], Chanyoung Jung[2]
Rohan Thakker[2], Nicholas Rhinehart[1], Ali-akbar Agha-mohammadi[2]
[*]Equal contribution [1]UC Berkeley, [2]NASA JPL

***Abstract*— Vision-based control of autonomous vehicles presents major challenges, particularly outside of well structured environments with clear road boundaries and lane markings. Learning control policies from human driving data offers an appealing alternative to classical navigation pipelines: by learning to directly associate observations with actions that avoid obstacles and achieve navigational goals, it is possible to circumvent many of the challenges associated with manually engineering a driving system for unstructured or off-road settings. However, integrating learning-based approaches into robust high-performance control systems presents a major challenge. In this paper, we describe a system for passenger-scale autonomous navigation in off-road environments that combines imitative models with low-level model-predictive control. Although the system learns to control the vehicle directly through perception, it is designed to integrate together learning-based components with constraints and trajectory optimization so as to provide a complete navigational system. Our experiments demonstrate the performance of the system in real-world scenarios over complex off-road terrains, and characterize its potential for improvement with the scaling of data collection and interventions. For a video description, see our link here**

Fig. 1: Left: full vehicle with sensor suite, which consists of a high-resolution stereo camera and depth sensor. Right: RGB and depth images as viewed from the onboard camera during data collection and real-time inference.

I. Introduction

A large amount of research and development in building large-scale off-road autonomy stacks has enabled implementations of off-road driving using a combination of environment mapping, path planning and model-based control [1], [2], [3]. Such a "geometric" stack tends to be very reliable by being conservative about its traversability estimates – through acting overly reliant on modeling the 3D geometry of the scene for path planning – and generally involves lots of hand-engineering (fine-tuning) and heuristics. Imitation learning (IL) [4], [5] is a powerful paradigm that can enable the system to learn relevant cues *directly* from prior experience and improve its performance as we gather more data. However, learning-based systems can be unreliable in the presence of out-of-distribution objects in the scene and difficult to integrate with existing autonomy stacks. An ideal autonomy solution would be able to leverage the ability of IL to learn directly from data, while not giving up the reliable conservativeness of a geometric stack.

In this work, we present **RACER-L4P**, the learning for planning (L4P) method for autonomous off-road driving that is enabled by data-driven learning from Deep Imitative Models (DIM) [6]. We train an IL-based planner to infer a 2D navigation path in top-down coordinates (rather than the conventionally used action predictions) and combine it with a model-based controller. While the particular components that we use to build **RACER-L4P** draw on prior work [6], [7], [8], our system serves as a proof-of-concept to enable passenger-sized autonomous driving over off-road terrain without any manual perceptual feature or cost-function engineering required as well as a novel instantiation of MPPI and IL. Our experiments demonstrate that **RACER-L4P** can successfully infer and follow acceptable paths in off-road unstructured environments. Furthermore, our system continually learns from incoming navigational data via collision intervention collection, creating a stack that is constantly improving itself in an iteration-based manner as shown in our performance analysis experiments.

II. Related Work

A. Current Off-Road Autonomy Stacks

Driving in off-road environments requires the vehicle to assess the traversability of the terrain. This can be achieved using geometry-based or appearance-based methods as summarized in Papadakis [9]. Geometry-based methods involve constructing a terrain map [10], [11] from depth measurements obtained via sensors such as LiDAR and stereo cameras. This terrain map is used to generate a traversability cost by performing stability analysis, using features like surface normals and the maximum or minimum height of the terrain, which can be used by motion planning and control algorithms to plan vehicle's actions [12], [13], [14]. Appearance-based methods incorporate higher-level costs through concepts such as semantic segmentation, object detection, and instance segmentation [15], [16], [17], where machine learning techniques are widely applied today. In contrast to these approaches, we describe a system that utilizes imitation learning to directly learn perception from

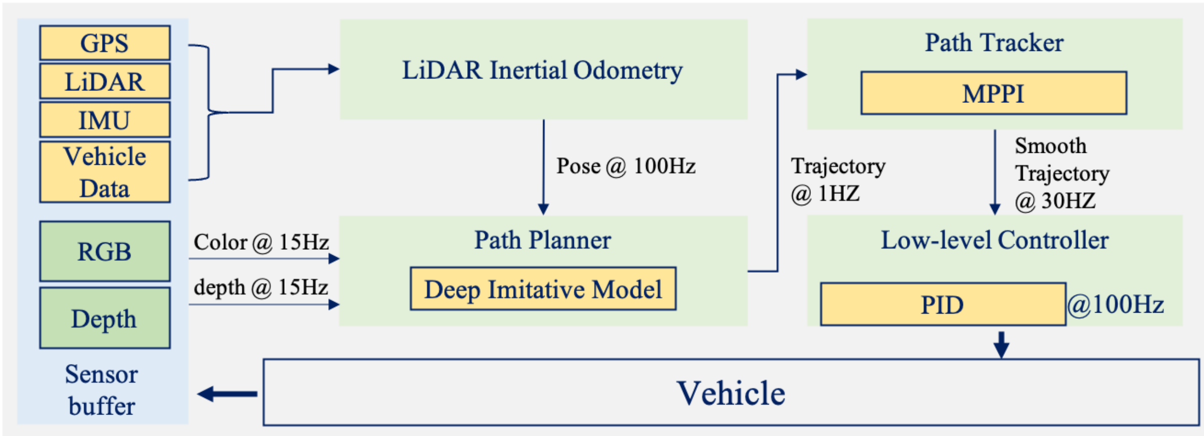

Fig. 2: Framework overview of our system. The system takes LiDAR-inertial odometry for past-trajectory input and RGB & depth images as visual inputs. It outputs the planned short-range path ($\leq$15m) using the DIM policy. The MPPI then smooths the path and generates position, velocity, throttle, and steer control commands to be applied to the low-level PID controller.

raw observations, bypassing an explicit representation of traversability across a map. We integrate this imitation learning process into a more conventional planning and control pipeline that still allows us to impose constraints and employ state-of-the-art tracking methods. This hybrid method is also robust as it scales with additional data collection whereas purely geometry-based methods cannot easily learn from incoming data automatically.

B. *Imitation Learning for Navigation and Driving*

Imitation Learning (IL) methods provide a simple way to implicitly model traversability by learning from prior experience. Expert demonstrations can provide a simple way to learn complex relationships between scene features and traversability. For navigating between points, goal-conditioned IL [18] has been used for control in many settings but often does not generalize well causing poor performance when out of distribution. Using IL with external goal direction has been shown to work well in simulation [6] and can be directly integrated with geometric information to improve out-of-distribution performance for off-road navigation [7]. But these algorithms have not been demonstrated on large vehicles in the real world. In [19], authors utilized an inverse reinforcement learning (IRL) framework for predicting traversability, which learns from human demonstrations, but was still required to repeatedly solve the MDP during inference which is not suitable for longer range of navigation and high-speed of driving.

Many similar on-road end-to-end methods for navigation have been explored for mapping raw input directly to steering angles [20], [21], [22], but these inherently differ by their problem formulation as they are on-road. Navigation on-road affords the model a largely unimodal distribution of potential paths and furthermore is heavily guided by road segmentation and line markers. Some attempts at using learning for end-to-end driving have been made [23], [24] but these methods do not use any explicit goal-direction, Instead, they do supervised behavior cloning (BC) to predict a most-likely target steer. More importantly, these methods only learn to implicitly drive fast while dodging obstacles, not navigate from point A to B. Our proposed method, on the other hand, works at the trajectory-level which enables heuristic-based scoring to follow high-level goals.

More explicit off-road traversability estimation methods that rely on learning exist and can classify full scenes [25] or semantically segment traversable areas [26], [1], but both require extensive annotation which can be hard to get for diverse, unstructured scenes in off-road environments. We propose a method for driving by implicitly learning traversability using unlabeled data.

III. Summary of the **RACER-L4P** System

A. *Hardware Setup and Sensor Suite*

We instantiate our **RACER-L4P** system on a Polaris MRZR-X platform (see Fig. 1), which offers the impressive capability to drive on a variety of challenging off-road terrains. The vehicle is equipped with a baseline sensor package, computing resources, and autonomy stack. The sensor package includes an inertial measurement unit (IMU), radar arrays, stereo camera pairs front and backward, and four lidars. The computing resources are comprised of networked GPUs populated with a basic "drive by wire" vehicle management system to execute driving commands (e.g., accelerate, break, turn). Our method uses the color and depth images from the RGB-D sensor package at the front of the vehicle.

B. *Software Stack*

The **RACER-L4P** software is built based on the NeBula system [8]. In this study, we are focusing on short range planning to achieve resilient navigation at high speeds in off-road settings. A traditional planner requires pre-built/fine-tuned prior information, such as occupancy information, traversability analysis, etc. As opposed to a traditional planner, our architecture requires only a short pose history of the vehicle, as well as an RGB-D image. Fig. 2 illustrates our software architecture.

Fig. 3: Example intervention scenes used for training. The deployed imitative policy often makes wrong predictions (plans shown in red), requiring a human expert to intervene and provide corrective behavior (shown in cyan). We use this corrective behavior as training data to continually improve our system.

IV. Learning Off-Road Navigation with Imitative Models

Our off-road driving software stack consists of two sub-systems: imitation-learning based spatial planning and control tracking. The purpose of the former is a data-driven method to infer the *highest-quality path given the perceptual data*, and the latter to infer the *controls that track this path*. For spatial planning, we adopt deep imitative models [6], which learn conditional density estimators of expert-like spatial trajectories, and use them to predict time-profiled trajectories at 10Hz for 4 seconds (i.e., each trajectory is 40 2D points). The planned trajectories are transmitted to the low-level controller to infer control actions. Specifically, we employ a Model Predictive Path Integral Controller (MPPI) [27] to generate command signals which can follow the trajectory precisely for our control sub-system.

A. Tracking with Model-Predictive Path Integral Control

We follow the MPPI framework to determine the controls to send to the robotic system when given a path to follow. MPPI works by optimizing a distribution that can be sampled for high-quality trajectories. After sampling a trajectory from the current distribution, it rolls out the dynamics with a predictive model. The control distribution desired is one which minimizes costs of control, state-dependent custom costs, and any constraints that the user may specify. To get the best distribution, we can treat the optimization problem as one of a KL-divergence minimization between a desired optimal distribution and the current trajectory distribution. This is done via iteratively updating the means of the current distribution with the rolled out dynamics.

B. Spatial Planning with Deep Imitation Learning

Geometric-based planning is a well-studied area in autonomous robotics. In general, geometric pipelines use costmaps built through traversability assessment using features based on the conventional algorithms such as surface fitting [28] and height based cost map generation [2]. However, rule-based algorithms like these are significantly affected by the quality of the parameters and often require significant computational resources for high-frequency collision checking and feature processing. More importantly, in rapidly changing environments, such as in off-road terrain, it is hard to perform robustly with static (no human-in-the-loop) geometric pipelines.

We now describe our proposed model's formulation and setup: let $o_t = (x_{\leq t}, i_t)$ denote the sensor information available to the vehicle, where $x_t \in \mathbb{R}^3$ is the vehicle's odometry position, and $i_t \in [0,1]^{H \times W \times 4}$ is an RGB-D image. Let $\tau \in \mathbb{R}^H$ denote a trajectory of potential future positions: $\tau \doteq x_{t+1:t+H}$, and τ_i denote x_{t+i}.

Our goal is to learn a cost-function for planning, C_{learned}. We use this function to select the best τ, τ^*. This trajectory is then given to the MPPI controller for tracking. Our receding-horizon path planner uses Eq. (1):

$$\tau^* = \operatorname{argmin}_\tau C_{\text{learned}}(\tau, o_{\leq t}) \tag{1}$$

C. Learned Imitative Cost

We design this learned cost to be a conditional probability density function of possible future trajectories approximated by a neural architecture. We learn this function, $q(\tau|o_{\leq t})$, by maximum likelihood estimation, following [6]. Once trained, we use the negative log-likelihood of a trajectory in the data-distribution as the learned cost function: $C_{\text{learned}}(\tau, o_{\leq t}) = -\log q(\tau|o_{\leq t})$. While [6] uses gradient-based planning to approximately identify τ^*, we found using a pre-generated library of paths to be more computationally efficient. We generate this library from the centroids of k-means on the training trajectories (we used $K = 200$).

D. Model Architecture

We design $q(\tau|o_{\leq t})$ to be a conditional autoregressive normalizing flow, which is a universal approximator (it can theoretically model any density function) [29]. A normalizing flow enables exact inference of an arbitrary point in event space, which allows us to evaluate the planning criterion exactly for any candidate trajectory. The model uses a learned function, f_{enc} to encode i_t and combines it with x_t via another learned function, g_{enc}. This creates a contextual vector, $z \in \mathbb{R}^d$ where d represents the contextual vector dimensionality. More precisely, we have $z = g_{enc}(f_{enc}(i_t), x_t)$ where f_{enc} is MobileNet-v2 [30] and g_{enc} is a Multilayer Perceptron. Given z and a base distribution sample $x \in \mathbb{R}^H$, we can generate a sample in the data-distribution, $\tau = F(x; o_{\leq t})$, $\tau \in \mathbb{R}^H$. Sampling from the model enables one to inspect the model's predictions for expert-like trajectories, although it is not used to plan (the model's probability density function is used to plan).

E. Distribution Shift

A well-known issue with offline imitation learning is the compounding errors problem [31]. More specifically, when an imitative model (or policy) that scores well on offline metrics is deployed, the visitation distribution can generally diverge from the training distribution – the deployment of the model violates a key supervised learning assumption of IID data. As this occurs, errors may begin building up and cause catastrophic failure due to poor performance of the function approximator when out of distribution. We address this problem by employing DAgger to improve deployment performance [31]: after initial training, we deployed the model and collected corrective maneuvers after interventions, then we combined the original training dataset with the interventions captured during deployment and retrained the model for redeployment. This DAgger intervention loop was then repeated multiple times. Example visualizations of this process are shown in Figure 3.

V. Evaluation on a Passenger-Sized Off-Road Vehicle

Our main metrics are based on the interventions by a human safety driver. When the deployed system starts behaving unsafely (e.g. may soon collide with an obstacle), the human safety driver intervenes to correct the vehicle. We recorded these interventions both as a measure of system performance and to further improve the system. We compute metrics based on the interventions per unit distance and interventions per unit time. We used 4 hours of driving data with 1 hour for evaluation offline.

A. Behavior Cloning Baseline

Furthermore, instead of modeling $q(\tau|o_{\leq t})$ as a conditional autoregressive normalizing flow, we also compare to modeling it as a conditional Gaussian distribution in trajectory space, which we refer to as the Behavior Cloning ("BC") model. A typical BC approach minimizes L2 error with a deterministic prediction, which is equivalent to fitting a Gaussian distribution and always outputting its mean. We expect a Gaussian distribution not to be able to realize the true multi-mode expert trajectory distribution, due to the high degrees of uncertainty generally present in driving, let alone off-road driving.

B. Experimental Results

We first study the performance of on-policy driving across model architecture with one iteration of intervention data. Specifically, we aim to analyze how much of an empirical performance boost we can get by using a conditional flow-based architecture instead of a BC model and how this scales with the inclusion of intervention data. Our results are shown in Table I and show how the flow model performs better after an iteration. As hypothesized, the obstacle rich nature of the trail motivates the usage of a multiple-mode distribution representation, which only the flow model can approximate.

We also sweep across data modality by running an ablation study across image data type. Our goal is to determine if, with the Flow-based architecture, the depth map is a significant component of the system setup. We do this by running identical experiments with RGB data and RGB-D data and show our results in Table II. Evidently, the inclusion of depth before any iterations seems to help performance.

Model	$\frac{\text{Interventions}}{\text{Minute}} \downarrow$	$\frac{\text{Interventions}}{\text{100m}} \downarrow$
BC, post-intervention data inclusion	2.479	2.833
Flow, post-intervention data inclusion	**2.004**	**2.197**

TABLE I: Compared to the BC model, it can be seen that the flow model shows about 1.2 times higher performance. This motivates the usage of the Flow for its better performance across 1 data aggregation iteration.

Model	$\frac{\text{Interventions}}{\text{Minute}} \downarrow$	$\frac{\text{Interventions}}{\text{100m}} \downarrow$
RGB-D Flow, pre-intervention data inclusion	**5.179**	**5.561**
RGB Flow, pre-intervention data inclusion	6.5	6.97

TABLE II: Online metrics with and without a concatenated depth map. In the flow model, utilizing a depth map is crucial, and informs how we should run our later dagger iterations.

For our final experiment, we use an RGB-D policy using a conditional-flow estimator as an imitative path-planner integrated with MPPI. We collect 3 DAgger iterations and show our results in Table III. As expected, the final iteration model showed improved performance by about 3 times.

Model Iteration	$\frac{\text{Interventions}}{\text{Minute}} \downarrow$	$\frac{\text{Interventions}}{\text{100m}} \downarrow$
Iteration 0	5.179	5.561
Iteration 1	2.004	2.197
Iteration 2	1.740	1.895
Iteration 3	**1.63**	**1.5**

TABLE III: Final DAgger iteration results are shown above for every iteration. After an iteration, the model was retrained with the interventions and re-evaluated. Inclusion of intervention data improves the model performance significantly and is scalable.

VI. Discussion

In this paper, we presented a system for learning-based visual navigation in off-road environments for passenger-sized vehicles. Our system combines an imitative model with trajectory optimization and model-predictive control to drive a vehicle through off-road, desert-like environments, and the design enables potential incorporation of other path-based costs, whether learning-based or hand-designed, as might come from a standard geometric pipeline.

Acknowledgment

The work is partially supported by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (80NM0018D0004), and Defense Advanced Research Projects Agency (DARPA).

References

[1] D. Maturana, P.-W. Chou, M. Uenoyama, and S. Scherer, "Real-time semantic mapping for autonomous off-road navigation," in *Proceedings of 11th International Conference on Field and Service Robotics (FSR '17)*, September 2017, pp. 335 – 350.

[2] J. Choi, J. Lee, D. Kim, G. Soprani, P. Cerri, A. Broggi, and K. Yi, "Environment-detection-and-mapping algorithm for autonomous driving in rural or off-road environment," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 2, pp. 974–982, 2012.

[3] K. Chu, M. Lee, and M. Sunwoo, "Local path planning for off-road autonomous driving with avoidance of static obstacles," *IEEE transactions on intelligent transportation systems*, vol. 13, no. 4, pp. 1599–1616, 2012.

[4] J. Ho and S. Ermon, "Generative adversarial imitation learning," *Advances in neural information processing systems*, vol. 29, 2016.

[5] X. B. Peng, A. Kanazawa, S. Toyer, P. Abbeel, and S. Levine, "Variational discriminator bottleneck: Improving imitation learning, inverse rl, and gans by constraining information flow," *arXiv preprint arXiv:1810.00821*, 2018.

[6] N. Rhinehart, R. McAllister, and S. Levine, "Deep imitative models for flexible inference, planning, and control," in *International Conference on Learning Representations*, 2020. [Online]. Available: https://openreview.net/forum?id=Skl4mRNYDr

[7] N. Dashora, D. Shin, D. Shah, H. Leopold, D. Fan, A. Agha-Mohammadi, N. Rhinehart, and S. Levine, "Hybrid imitative planning with geometric and predictive costs in off-road environments," in *2022 International Conference on Robotics and Automation (ICRA)*, 2022, pp. 4452–4458.

[8] A. Agha, K. Otsu, B. Morrell, D. D. Fan, R. Thakker, A. Santamaria-Navarro, S.-K. Kim, A. Bouman, X. Lei, J. Edlund *et al.*, "Nebula: Quest for robotic autonomy in challenging environments; team costar at the darpa subterranean challenge," *arXiv preprint arXiv:2103.11470*, 2021.

[9] P. Papadakis, "Terrain traversability analysis methods for unmanned ground vehicles: A survey," *Engineering Applications of Artificial Intelligence*, vol. 26, no. 4, pp. 1373–1385, 2013. [Online]. Available: http://dx.doi.org/10.1016/j.engappai.2013.01.006

[10] P. Fankhauser, M. Bloesch, C. Gehring, M. Hutter, and R. Siegwart, "Robot-centric elevation mapping with uncertainty estimates," in *International Conference on Climbing and Walking Robots (CLAWAR)*, 2014.

[11] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, "Octomap: An efficient probabilistic 3d mapping framework based on octrees," *Autonomous robots*, vol. 34, no. 3, pp. 189–206, 2013.

[12] K. Otsu, G. Matheron, S. Ghosh, O. Toupet, and M. Ono, "Fast approximate clearance evaluation for rovers with articulated suspension systems," *Journal of Field Robotics*, vol. 37, no. 5, pp. 768–785, 2019.

[13] R. Thakker, N. Alatur, D. D. Fan, J. Tordesillas, M. Paton, K. Otsu, O. Toupet, and A.-a. Agha-mohammadi, "Autonomous off-road navigation over extreme terrains with perceptually-challenging conditions," in *International Symposium on Experimental Robotics*. Springer, 2020, pp. 161–173.

[14] P. Krüsi, P. Furgale, M. Bosse, and R. Siegwart, "Driving on point clouds: Motion planning, trajectory optimization, and terrain assessment in generic nonplanar environments," *Journal of Field Robotics*, vol. 34, no. 5, pp. 940–984, 2017.

[15] M. J. Procopio, J. Mulligan, and G. Grudic, "Learning terrain segmentation with classifier ensembles for autonomous robot navigation in unstructured environments," *Journal of Field Robotics*, vol. 26, no. 2, pp. 145–175, 2009.

[16] H. Lu, L. Jiang, and A. Zell, "Long range traversable region detection based on superpixels clustering for mobile robots," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 546–552.

[17] P. Ewen, A. Li, Y. Chen, S. Hong, and R. Vasudevan, "These maps are made for walking: Real-time terrain property estimation for mobile robots," *IEEE Robotics and Automation Letters*, 2022.

[18] Y. Ding, C. Florensa, M. Phielipp, and P. Abbeel, "Goal-conditioned imitation learning," *Advances in Neural Information Processing Systems*, 2019. [Online]. Available: http://arxiv.org/abs/1906.05838

[19] Y. Zhang, W. Wang, R. Bonatti, D. Maturana, and S. Scherer, "Integrating kinematics and environment context into deep inverse reinforcement learning for predicting off-road vehicle trajectories," *arXiv preprint arXiv:1810.07225*, 2018.

[20] D. A. Pomerleau, "Alvinn: An autonomous land vehicle in a neural network," in *NIPS*, 1988.

[21] M. Bojarski, D. Del 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," 2016. [Online]. Available: https://arxiv.org/abs/1604.07316

[22] M. Bojarski, P. Yeres, A. Choromanska, K. Choromanski, B. Firner, L. Jackel, and U. Muller, "Explaining how a deep neural network trained with end-to-end learning steers a car," 2017. [Online]. Available: https://arxiv.org/abs/1704.07911

[23] Y. Pan, C.-A. Cheng, K. Saigol, K. Lee, X. Yan, E. A. Theodorou, and B. Boots, "Agile autonomous driving using end-to-end deep imitation learning," *Robotics: Science and Systems XIV*, 2018.

[24] Y. LeCun, U. Muller, J. Ben, E. Cosatto, and B. Flepp, "Off-road obstacle avoidance through end-to-end learning," in *NIPS*, 2005.

[25] I. Kostavelis, L. Nalpantidis, and A. Gasteratos, "Supervised traversability learning for robot navigation," in *Towards Autonomous Robotic Systems*, R. Groß, L. Alboul, C. Melhuish, M. Witkowski, T. J. Prescott, and J. Penders, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 289–298.

[26] A. Shaban, X. Meng, J. Lee, B. Boots, and D. Fox, "Semantic terrain classification for off-road autonomous driving," in *5th Annual Conference on Robot Learning*, 2021. [Online]. Available: https://openreview.net/forum?id=AL4FPs84YdQ

[27] A. A. Grady Williams and E. A. Theodorou, "Model predictive path integral control: From theory to parallel computation," *Journal of guidance, control and dynamics (JGCD)*, 2017.

[28] A. Broggi, E. Cardarelli, S. Cattani, and M. Sabbatelli, "Terrain mapping for off-road autonomous ground vehicles using rational b-spline surfaces and stereo vision," in *2013 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2013, pp. 648–653.

[29] G. Papamakarios, E. T. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, "Normalizing flows for probabilistic modeling and inference." *J. Mach. Learn. Res.*, vol. 22, no. 57, pp. 1–64, 2021.

[30] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[31] S. Ross, G. Gordon, and D. Bagnell, "A reduction of imitation learning and structured prediction to no-regret online learning," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2011, pp. 627–635.