



HAL
open science

Evaluating Robustness over High Level Driving Instruction for Autonomous Driving

Florence Carton, David Filliat, Jaonary Rabarisoa, Quoc Cuong Pham

► **To cite this version:**

Florence Carton, David Filliat, Jaonary Rabarisoa, Quoc Cuong Pham. Evaluating Robustness over High Level Driving Instruction for Autonomous Driving. IV 2021 - 32nd IEEE Intelligent Vehicles Symposium, Jul 2021, Nagoya, Japan. hal-03377799

HAL Id: hal-03377799

<https://hal.science/hal-03377799v1>

Submitted on 14 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evaluating Robustness over High Level Driving Instruction for Autonomous Driving

Florence Carton^{1,2}, David Filliat¹, Jaonary Rabarisoa² and Quoc Cuong Pham²

Abstract—In recent years, we have witnessed increasingly high performance in the field of autonomous end-to-end driving. In particular, more and more research is being done on driving in urban environments, where the car has to follow high level commands to navigate. However, few evaluations are made on the ability of these agents to react in an unexpected situation. Specifically, no evaluations are conducted on the robustness of driving agents in the event of a bad high-level command. We propose here an evaluation method, namely a benchmark that allows to assess the robustness of an agent, and to appreciate its understanding of the environment through its ability to keep a safe behavior, regardless of the instruction.

I. INTRODUCTION

Autonomous driving can be achieved using two main paradigms: modular pipeline and end-to-end driving. In the case of the modular pipeline, driving is divided into several sub-modules such as: detection (of road lines, other vehicles, pedestrians, etc...), path planning, and control to compute the actions to be performed (accelerate / brake, steer). This approach allows to train each sub-modules separately and therefore offers a finer control on the agent behavior and a more interpretable training. On the other hand, optimizing sub-modules separately can lead to a sub-optimal policy, since all modules are independent. In the case of end-to-end driving, the commands are calculated directly from the inputs, allowing direct optimization, and no intermediate steps.

A lot of recent approaches rely on end-to-end autonomous driving whether through imitation learning [1], [2], [3], [4], [5], [6] or reinforcement learning [7], [8], [9], [10]. In imitation learning the agent learns by imitating expert’s behavior, whereas in reinforcement learning the agent learns by trial and error, based on a reward signal that indicates the correctness of its behavior.

While early work on end-to-end driving initially focused on the task of road following, many are now concentrating on driving in urban areas, which involves intersection management. To manage an intersection, the



Fig. 1. What to do if the GPS breaks down?

agent must take into account a high-level command (e.g. turn left) provided by a high-level planning system to reach its final goal. Nevertheless, the agent must be robust in the sense that the information in the environment must keep priority over the high-level order. This can be for example the case where the gps system indicates to turn, but the street is blocked for road works, or the case where the gps signal is lost due to bad connection and it does not detect that the vehicle is approaching an intersection (Figure 1). While enforcing such priority is natural in the modular approach, it is not the case for end-to-end approaches and we should therefore evaluate if the controller follows this priority.

In order to highlight the importance of robust decision making of the driving agent, we propose a new way to evaluate their behavior, based on their reaction to mishap. More precisely, we evaluate the stability of an agent over driving instruction, by proposing a benchmark that will evaluate the reaction of the agent if a wrong command is given. We believe these tests are a basis for a necessary evaluation of autonomous driving systems safety.

We then evaluate several state of the art approaches in this new benchmark, and compare the effect of different training approaches on an agent robustness.

¹ U2IS, ENSTA Paris, INRIA FLOWERS, Institut Polytechnique de Paris, Palaiseau, France florence.carton@ensta.fr, david.filliat@ensta.fr

² Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France jaonary.rabarisoa@cea.fr, quoc-cuong.pham@cea.fr

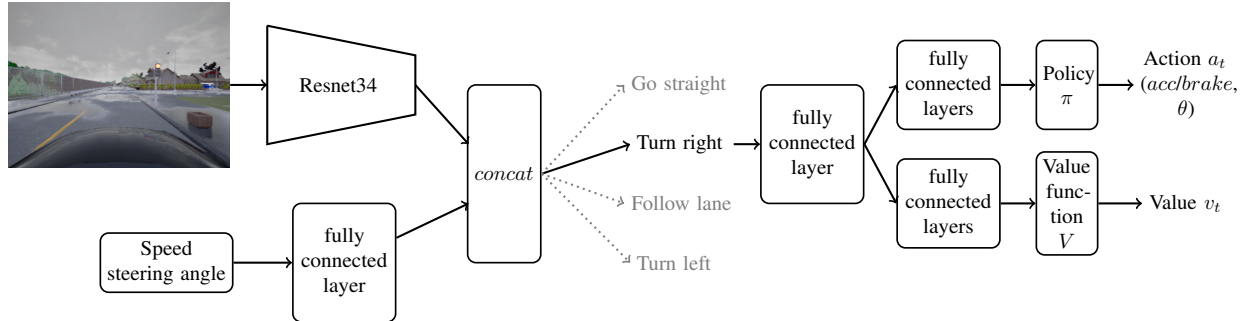


Fig. 2. Reinforcement learning architecture : model with branch

II. RELATED WORK

Autonomous driving is a field that has been expanding rapidly in recent years. Since Alvin [11] in the 90s, the performance of autonomous cars based on visual inputs has been steadily increasing. Many approaches use supervised learning [1], [11], [2], where the agent learns to imitate the behavior of an expert. However we can observe an increasing interest for autonomous driving using reinforcement learning [9], [10], [8], [12].

But whereas a lot of work only handle lane following task [9], [1], [13], more and more work is being done on the task of driving in urban environments [2], [14], [8], [10], [15]. With the introduction of Carla simulator [16], which proposes a number of urban environments, training autonomous driving in the city can be carried out safely in simulation. Different methods are used to indicate to the agent the route to follow. Some use the map and the route to follow (in birdview) [17], other use the waypoints [18] or intermediate goals [3], but this requires access to environmental information that is not always available. Finally, one possibility is to give high level commands (straight ahead, turn right, etc...). To take into account high level control, a branch architecture producing as many output as available high-level commands (Figure 2) is often used [16], [2], [8]. It is more effective in forcing the agent to take into account the command [2] than a simple concatenation of the command to the neural network input [19].

Two approaches have recently achieved (almost) perfect score on Carla CoRL Benchmark [2]. The first one with imitation learning named *Learning by Cheating* [15] consists in a two stages training. First a privileged agent that has access to the layout of the environment is trained with expert demonstrations. This privileged agent is then used as an expert to train in a supervised way an agent who only has raw image input. The second approach called MaRLn [8], designed by a team from Valeo, is based on reinforcement learning and affordance. A network is first trained in a supervised way to segment the road and predict information about the environment

(distance to next crossing, presence of traffic light, state of this traffic light...). Then the intermediate representations obtained are used as input for a RL agent. This approach reached high performance on the Carla challenge [20], [21].

As we can see, recent years have witnessed a strong increase in the performance of end-to-end driving, however, as far as we know, no studies have been conducted on the robustness of these approaches. The generalization of driving agents is classically evaluated, by evaluating them in cities that have not been seen, with weather and luminosity conditions different from training ones [16], or dense dynamic agents [22]. Some have focused on the robustness of driving agents, but usually only deal with the robustness of the detection part [23], [24]. In *Improved Robustness and Safety for Autonomous Vehicle Control with Adversarial Reinforcement Learning*, Ma et al. [25] propose a framework to train a robust agent with reinforcement learning, and evaluate it by applying disturbance. The disturbances are applied on acceleration and steering angle. In the same way in ChauffeurNet [26], the agent is trained to be able to handle a trajectory disturbance, and in the Carla Challenge [20], [21], one of the scenarios simulates a loss of control of the vehicle - due to a very wet road. However, no evaluation is made on the reaction of the agents if a wrong command is given. We propose here a benchmark that would allow an evaluation of the ability of these agents to react in case of unexpected command, and thus measure their ability to understand the environment around them in the sense that they should keep producing a safe control corresponding to the environment situation. We will compare the performance of three of our agents trained using different approaches with two from the state of the art [15] and [8] that achieve high or even perfect scores on Carla CoRL benchmark [16].

III. EVALUATED AGENTS

In this section we describe our driving agents. We trained three different agents, all with reinforcement

learning. The first one is trained with semantic segmentation as input (as a baseline reaching perfect score on the benchmark), the second with RGB images, and the last one also inputs RGB images but uses auxiliary task to predicts semantic segmentation. Our agents are trained on the version 0.9.6 of Carla simulator [16] and are all evaluated on the original Carla CoRL benchmark with navigation task. For the purpose of this paper, we do not evaluate the performance on dynamic navigation (i.e. with other dynamic agents - cars and pedestrians) since we only want to evaluate the robustness on wrong commands.

A. Model and RL training

We have chosen the PPO algorithm [27], which has proven its effectiveness with input images and continuous actions. PPO belongs to the actor-critic category, and therefore calculates both policy and value function. We use the PPO implementation from [28].

1) *Architecture*: Among our 3 agents, the general architecture remains the same (Figure 2). The input (RGB image or semantic segmentation) is passed through an encoder (Resnet34 in the case of an image, and the smaller NatureCNN from [29] for the semantic segmentation). RGB image inputs are subject to standard data augmentation (partial erasing, Gaussian blur, variation of hue, saturation, brightness and saturation). The encoder is trained from scratch for the agents with segmentation as input and the one using auxiliary task, and pretrained with semantic segmentation task for the third one. Current speed and steering angle, which are passed through fully connected layers, are also used as input. The features are then concatenated and the output of the network is composed of the action and the value. To take into account the high-level instruction, a model with branches is used: the action and value outputs are duplicated for each high-level command, and the output of the corresponding branch is used depending on the command issued to the agent. For the agent using the auxiliary task, a decoder outputting semantic segmentation is added to

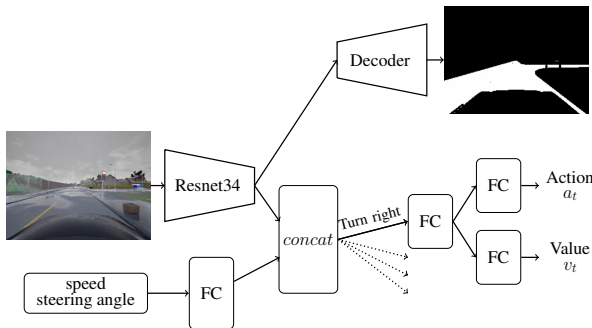


Fig. 3. Reinforcement learning architecture with auxiliary task

the model (see Figure 3). The decoder is inspired from Linknet [30], which is a segmentation network based on Resnet34. The auxiliary task is trained at the same time as driving task, 3 blocks of the Resnet34 encoder are common to segmentation and driving, the last block is duplicated and task-specific. Training is carried out with a homoscedastic combination of losses (supervised cross entropy loss for segmentation, and reinforcement loss for driving) from [31]. Equation 1 shows the loss function with l_{rl} and l_{aux} respectively the RL loss and auxiliary loss, and σ_{rl} and σ_{aux} two trainable parameters.

$$l = e^{-\sigma_{rl}} \times l_{rl} + \sigma_{rl} + e^{-\sigma_{aux}} \times l_{aux} + \sigma_{aux}. \quad (1)$$

2) *Reward function*: As the reward is the only information the agent obtains from the environment, it is a very important component in reinforcement learning algorithms. The reward function used is a weighted sum of three terms :

$$r = w_s \times r_{speed} + w_{cte} \times r_{cte} + w_a \times r_{angle} \quad (2)$$

with

- $r_{speed} = target_speed - |target_speed - speed|$: Computes the gap between target speed and current speed. Target speed is 35 km/h in general, and 15 km/h at intersections.
- $r_{cte} = (1 - d)$ with d the cross-track error, i.e. the lateral distance (in meters) of our agent to road center (Figure 4)
- $r_{angle} = (15 - |\alpha|)$ with α the angle (in degree) between the agent's direction and the road direction (Figure 4). Constants are added in r_{cte} and r_{angle} so that the corresponding reward is positive if the car is roughly aligned with the road (ie $\alpha < 15^\circ$) and not too far from road center ($d < 1m$)
- w_s , w_{cte} and w_a their associated weights, respectively 1.0, 10 and 0.1. The weights for every component were determined experimentally.

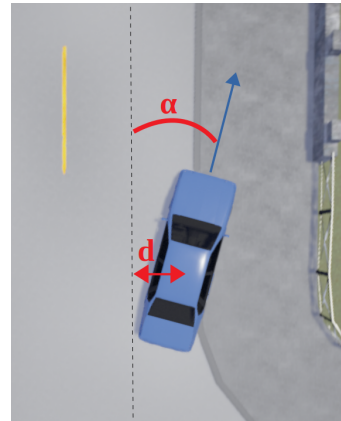


Fig. 4. Illustration of cross-track error d and angle α used to compute the reward

Training	Training conditions	New weather	New town	New Town & weather
LBC [15]	100%	100%	98%	100%
MaRLn [8]	100%	100%	100%	98%
Segmentation input	100%	100%	100%	100%
RGB image input	82%	98%	49%	40%
Auxiliary task	90%	92%	78%	68%

TABLE I

RESULTS OF OUR AGENTS AND SOTA AGENT ON CoRL BENCHMARK WITH NAVIGATION TASK

B. Results on CoRL Benchmark

CoRL Benchmark [32] is a goal directed urban navigation benchmark. Four different tasks can be evaluated (from going straight to full navigation with dynamic obstacles), and we will focus on navigation task (navigation without other cars or pedestrians). CoRL benchmark evaluates the generalization capabilities of the agent, by testing training conditions as well as unseen town and new weather conditions. The benchmark outputs the percentage of successful episodes, i.e. when the driving agent has managed to reach destination. For every episode, the trajectory is computed by the simulator and high level commands are sent to the agent with the different information (road picture, current speed, etc...). In the CoRL benchmark, given commands are always correct.

We present in Table I the performance of our agents, and compare them to the state of the art algorithms *Learning by Cheating* (LBC) [15] and MaRLn [8] on the navigation task. We used the open-source implementations of the benchmark released by *Learning by Cheating*¹ and *MaRLn*². For the latter, in order to make a fair comparison, we used the agent who was trained only on training conditions for the CoRL benchmark. The score is the percentage of success of requested trajectories. Evaluation is made with the version 0.9.6 of Carla.

We find that among our agents, the one which competes with the state of the art is the one which uses semantic segmentation as input. When using RGB images as input, the performances fall especially in the city that was not seen during the training, although the use of the auxiliary task allows for better generalization. The lower results of our agents with image input in training conditions compared to new weathers are due to a harder weather (hard rain) among training weathers compared to the test weathers. However, the results of our agents with the input image remain high in the training city, with more than 90% successful episodes on training town for our agent with auxiliary task.

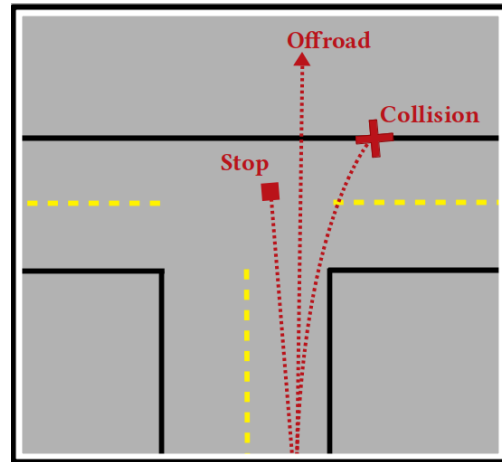


Fig. 5. Examples of unsafe behaviors.

IV. WRONG COMMAND BENCHMARK

Usually in driving benchmark, tests are made by changing the visual input : different towns and weathers that were not seen during training. This allows to check the generalization capability of the agent. However in autonomous driving, there are not only visual inputs. In this section, we want to measure the stability of driving agents over given instructions.

A. Benchmark Description

Our benchmark is inspired by the CoRL benchmark from Carla [16]. It consists in giving an incorrect command at an intersection, and evaluating the reaction of the driving agent. Since we only work in Town01, there are 4 possible commands : follow lane (basic command when there is no crossing), go straight, turn left and turn right. We want to evaluate the reaction of the driving agent if an incorrect command is given at an intersection. We can distinguish two cases : either the agent is told to go at a direction that does not exist, or the basic command (follow lane) is still given even if the agent reaches a crossing. We will measure the number of unsafe behavior (i.e. collision, offroad, or stop at the middle of

¹<https://github.com/dotchen/LearningByCheating>

²<https://github.com/valeoai/LearningByCheating>

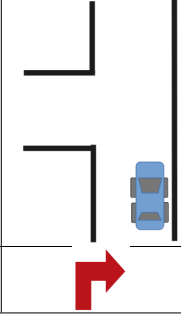
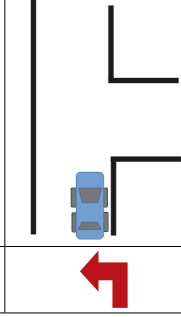
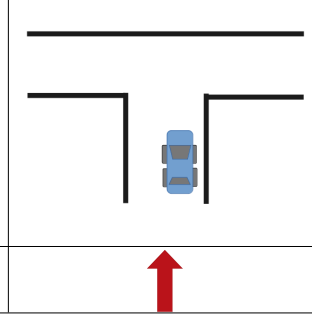



Type of crossing					
	wrong command				Total unsafe
LBC [15]	Unsafe behavior with wrong command	0	5	10	23 / 60
	Unsafe behavior with 'follow lane' command	0	0	8	
MaRLn [8]	Unsafe behavior with wrong command	2	1	10	13 / 60
	Unsafe behavior with 'follow lane' command	0	0	0	
RL seg input	Unsafe behavior with wrong command	0	10	9	35 / 60
	Unsafe behavior with 'follow lane' command	0	6	10	
RL img input	Unsafe behavior with wrong command	0	1	7	16 / 60
	Unsafe behavior with 'follow lane' command	0	4	4	
RL with auxilliary task	Unsafe behavior with wrong command	0	2	0	3 / 60
	Unsafe behavior with 'follow lane' command	1	0	0	

TABLE II

RESULTS OF OUR AGENTS AND TWO STATE OF THE ART AGENTS ON OUR WRONG-COMMAND BENCHMARK, SHOWING THE NUMBER OF UNSAFE BEHAVIOR ON 10 TRIALS.

the crossing) in these cases (see Figure 5 for examples of unsafe behaviors). This will help to measure if the agent has achieved a certain understanding of the environment, or if it blindly follows instructions. For each crossing type, we took 10 of them in the training town, with the ClearNoon time/weather preset, a simple and already seen during training weather condition, and run small episodes starting just before the crossing with erroneous command.

In the event that a wrong command is given at a junction, one can expect two behaviors from a human driver: either it stops before the junction or it randomly takes one of the proposed routes. Out of the 10 test episodes, we have therefore counted as unsafe behavior the cases where the agent stops in the middle of the crossing, and the cases where the agent collides or goes offroad. For every crossing type we counted the unsafe behavior for both wrong command (that is not available at the given crossing), and basis - follow lane - command. The results are given in table II for state of the art algorithm [15] and [8] as well as for our three agents: with the segmentation and with the image in input, and the last one with the segmentation as auxiliary task. Since

the results measure unsafe behavior here, the smaller the more robust. The companion video illustrates some of the agents behaviors³. The code for the benchmark as well as the weights of our RL agent with auxiliary task can be found open-source⁴.

B. Qualitative Analysis

We first note that the results are very disparate according to the type of crossing. For the first crossing type, where the 'turn right' command is not possible, all the agents drive correctly (with sometimes bites into the sidewalk, but overall safe behaviors). This is probably due to the fact that the sidewalk is always present to the right of the agent, and that this helps the agent to drive correctly.

For the second crossing, where the 'turn left' command is not possible, the agents have different behaviors. The LBC [15] agent deviates slightly if it is still asked to turn left, and then in half of the cases, stops abruptly in

³<https://github.com/CEA-LIST/AD-RobustnessEval/blob/master/video/Demo.mp4>

⁴<https://github.com/CEA-LIST/AD-RobustnessEval/>

the middle of the crossing, which we considered to be dangerous and therefore unsafe behavior. The MaRLn [8] agent keeps its course, and both [8] and [15] perform well when the command remains to follow the road. Among our agents, the behavior varies greatly. The agent using input segmentation has a very high collision rate, because it turns directly left if this command is given. The two agents using the input image react better, even though the auxiliary task seems to help to behave correctly.

Finally the third type of intersection, where it is not possible to go straight, appears to be the more difficult for all agents. The LBC agent has a very bad error management : in the vast majority of cases, it keeps going straight, even if there is a wall, and usually hit any obstacle in front of the car, and so does our agent that uses segmentation. The MaRLn agent manages to turn when the command follow lane is still given - by sometimes biting into the other line, but keeps going straight if the command straight is given. However, unlike LBC, it rarely collides, preferring to stop in the middle of the crossing. On the other hand, our agent which has learned with an auxiliary task does not make any mistakes. It turns arbitrarily to one side or the other, even if it is asked to go straight, which shows that it relies much more on the analysis of the environment around.

C. Discussion

We present the performance of 5 driving agents: one trained with imitation learning (LBC [15]), and the four others with reinforcement learning. We are not surprised by the rather low performance in terms of robustness of LBC. This is one of the drawbacks of supervised learning, since the cases of error are not present in the learning dataset, the agent never learns how to react to unexpected situations.

One might have thought that reinforcement learning, because of its trial-and-error nature, would alone have increased the robustness of learning. However, the disparity in the results of agents trained through reinforcement learning shows that this is not enough. We can compare the performance of our two agents with RGB image as input. Although using an encoder pretrained on the segmentation task, the simple agent (i.e. without auxiliary task) is less robust than the one using segmentation as an auxiliary task. The use of semantic segmentation as auxiliary task allows the agent to make the link between the semantic parts of the image and the rewards, and thus to have a better understanding of the environment. It is able to analyze its surroundings, and to make a decision in a situation it has never seen before by giving priority to the visual information over the high-level command.

MaRLn agent [8] is quite close to ours using auxiliary task, with the following differences : the encoder is pretrained to calculate segmentation as well as other

affordances (distance to traffic lights, position in the road...), and frozen during RL driving training. Perception is then decoupled from driving. Our agent on the other hand is trained end-to-end with both segmentation and driving. We believe that decoupling perception and control may reduce the robustness of the agent.

We have also analyzed the different experiments of our agent using only segmentation, and a possible explanation for the poor performance in terms of robustness is that it is too confident on the road marking perception. Indeed, the segmentation in Carla being perfect, this agent has never needed to use anything other than the road lines to learn to drive. In normal situations, when it reaches a crossroads, our agent starts to turn in the prescribed direction, and adjusts its trajectory with the lines of the road it joins (road markings are not present in the intersection). However, in the case of a wrong command, it starts to turn and cannot correct itself because there are no more lines - because the target road doesn't exist.

On the other hand, the agents which learned with an RGB image as input were subject to data augmentation. In particular, road lines are not always visible in the image (very rainy weather or Gaussian blur), so the agents learn to use other elements - the sidewalk for example. These agents therefore learn with information redundancy, which is not the case if only perfect segmentation is used. One lead for future work to validate this hypothesis would be to train an agent with segmentation input, by regularly and randomly removing one or more masks. The agent would then learn to use the different elements present in the observation in a robust way.

V. CONCLUSION

In this paper we propose a first study to evaluate the robustness to wrong command of a vision based driving agent. With the development of end-to-end conditional autonomous vehicles, we believe this issue is of crucial interest and has not been dealt with so far. We found in this evaluation that an agent appears to be more robust when perception and control are not decoupled during training. We also observed a better robustness when the agent has a high-level understanding of its environment with, for example, the ability to segment the observation, but also when it learns to use the redundancy of the elements present in the observation.

As future work, We can imagine other tests to evaluate the robustness: changing the visual input by placing the camera at a different position, changing steering wheel hardness, etc... Since all the situations cannot be seen during training, we need to have agents that are capable of adaption and understanding of the environment, in the sense that they must extract enough information to avoid dangerous situations.

REFERENCES

- [1] 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," pp. 1–9, 2016. [Online]. Available: <http://arxiv.org/abs/1604.07316>
- [2] F. Codevilla, M. Müller, A. López, V. Koltun, and A. Dosovitskiy, "End-to-end Driving via Conditional Imitation Learning," *Proceedings ICRA, 2018*, 2017.
- [3] N. Rhinehart, R. McAllister, and S. Levine, "Deep Imitative Models for Flexible Inference, Planning, and Control," *International Conference on Learning Representations*, no. 1, pp. 1–20, 2020. [Online]. Available: <http://arxiv.org/abs/1810.06544>
- [4] J. Hawke, R. Shen, C. Gurau, S. Sharma, D. Reda, N. Nikolov, P. Mazur, S. Micklethwaite, N. Griffiths, A. Shah, and A. Kendall, "Urban Driving with Conditional Imitation Learning," 2019. [Online]. Available: <http://arxiv.org/abs/1912.00177>
- [5] M. Toromanoff, E. Wirbel, C. Vejarano, X. Perrotton, and F. Moutarde, "End to End Vehicle Lateral Control Using a Single Fisheye Camera," pp. 3613–3619, 2018. [Online]. Available: <http://arxiv.org/abs/1808.06940>
- [6] Z. Yang, Y. Zhang, J. Yu, J. Cai, and 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.
- [7] E. Perot, M. Jaritz, M. Toromanoff, and R. D. Charette, "End-to-End Driving in a Realistic Racing Game with Deep Reinforcement Learning," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol. 2017-July, pp. 474–475, 2017.
- [8] M. Toromanoff, E. Wirbel, and F. Moutarde, "End-to-End Model-Free Reinforcement Learning for Urban Driving using Implicit Affordances," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7153–7162. [Online]. Available: <http://arxiv.org/abs/1911.10868>
- [9] A. Kendall, J. Hawke, D. Janz, P. Mazur, D. Reda, J.-M. Allen, V.-D. Lam, A. Bewley, and A. Shah, "Learning to Drive in a Day," 2018. [Online]. Available: <http://arxiv.org/abs/1807.00412>
- [10] X. Liang, T. Wang, L. Yang, and E. Xing, "CIRL: Controllable Imitative Reinforcement Learning for Vision-based Self-driving," 2018. [Online]. Available: <http://arxiv.org/abs/1807.03776>
- [11] D. A. Pomerleau, "ALVINN: an autonomous land vehicle in a neural network," pp. 305–313, 1989. [Online]. Available: <http://dl.acm.org/citation.cfm?id=89851.89891>
- [12] T. Agarwal, H. Arora, T. Parhar, S. Deshpande, and J. Schneider, "Learning to Drive using Waypoints," in *Machine Learning for Autonomous Driving Workshop at the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*, 2019.
- [13] P. Wolf, C. Hubschneider, M. Weber, A. Bauer, J. Hartl, F. Durr, and J. M. Zollner, "Learning how to drive in a real world simulation with deep Q-Networks," in *IEEE Intelligent Vehicles Symposium, Proceedings*, no. Iv, 2017, pp. 244–250.
- [14] A. Sauer, N. Savinov, and A. Geiger, "Conditional Affordance Learning for Driving in Urban Environments," no. CoRL, pp. 1–16, 2018. [Online]. Available: <http://arxiv.org/abs/1806.06498>
- [15] D. Chen, Z. Brady, K. Vladlen, and K. Hilipp, "Learning by Cheating," *CoRL*, no. CoRL, pp. 1–10, 2019. [Online]. Available: <http://vladlen.info/publications/learning-by-cheating/>
- [16] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An Open Urban Driving Simulator," *1st Conference on Robot Learning (CoRL 2017)*, pp. 1–16, 2017. [Online]. Available: <http://arxiv.org/abs/1711.03938>
- [17] J. Chen, B. Yuan, and M. Tomizuka, "Model-free Deep Reinforcement Learning for Urban Autonomous Driving," in *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 2765–2771. [Online]. Available: <http://arxiv.org/abs/1904.09503>
- [18] R. Agarwal, C. Liang, D. Schuurmans, and M. Norouzi, "Learning to Generalize from Sparse and Underspecified Rewards," 2019. [Online]. Available: <http://arxiv.org/abs/1902.07198>
- [19] C. Hubschneider, A. Bauer, M. Weber, and J. M. Zollner, "Adding navigation to the equation: Turning decisions for end-to-end vehicle control," *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, vol. 2018-March, pp. 1–8, 2018.
- [20] "Carla autonomous driving challenge 2019," <https://carlachallenge.org/>, 2019.
- [21] "Carla autonomous driving challenge 2020," <https://leaderboard.carla.org/challenge/>, 2020.
- [22] F. Codevilla, E. Santana, A. M. López, and A. Gaidon, "Exploring the Limitations of Behavior Cloning for Autonomous Driving," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 9329–9338, 2019. [Online]. Available: <http://arxiv.org/abs/1904.08980>
- [23] W. Zhou, J. S. Berrio, S. Worrall, and E. Nebot, "Automated evaluation of semantic segmentation robustness for autonomous driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 5, pp. 1951–1963, 2019.
- [24] C. Michaelis, B. Mitzkus, R. Geirhos, E. Rusak, O. Bringmann, A. S. Eckery, M. Bethgey, and W. Brendely, "Benchmarking Robustness in Object Detection: Autonomous driving when winter is coming," *arXiv preprint*, 2019.
- [25] X. Ma, K. Driggs-Campbell, and M. J. Kochenderfer, "Improved Robustness and Safety for Autonomous Vehicle Control with Adversarial Reinforcement Learning," *arXiv*, no. Iv, pp. 1665–1671, 2019.
- [26] M. Bansal, A. Krizhevsky, and A. Ogale, "ChauffeurNet: Learning to Drive by Imitating the Best and Synthesizing the Worst," pp. 1–20, 2018. [Online]. Available: <http://arxiv.org/abs/1812.03079>
- [27] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," pp. 1–12, 2017. [Online]. Available: <http://arxiv.org/abs/1707.06347>
- [28] A. Hill, A. Raffin, M. Ernestus, A. Gleave, A. Kanervisto, R. Traore, P. Dhariwal, C. Hesse, O. Klimov, A. Nichol, M. Plappert, A. Radford, J. Schulman, S. Sidor, and Y. Wu, "Stable baselines," <https://github.com/hill-a/stable-baselines>, 2018.
- [29] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning." *Nature*, vol. 518, no. 7540, pp. 529–33, 2015. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/25719670>
- [30] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," *2017 IEEE Visual Communications and Image Processing, VCIP 2017*, vol. 2018-Janua, pp. 1–4, 2018.
- [31] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" *Advances in Neural Information Processing Systems*, vol. 2017-Decem, no. Nips, pp. 5575–5585, 2017.
- [32] A. Dosovitskiy and V. Koltun, "Learning to Act by Predicting the Future," *ICLR 2017*, pp. 1–14, 2017.