Vilnius University

An investigation of deep imitation learning for mobile robot navigation

Shubham Juneja
Supervisor: Dr. Virginijus Marcinkevičius
Semester 2

Vilnius University

Plan of studies & implementation summary

Study year	Exams		Conference participations		Publications	
	Planned	Completed	Planned	Completed	Planned	Completed
I (2020/2021)	2	2	0	1		
II (2021/2022)	2	0				
III (2022/2023)			0	0	1	0
IV (2023/2024)			1	0	1	0

Report of activity plan

Exams		Conference Participation		Publications	
Planned	**Status**	**Planned**	**Status**	**Planned**	**Status**
Machine Learning	Passed with score of 9/10	All Sensors 2021, Nice, France	Paper accepted and presented at All sensors 2021 conference at Nice, France. On the 20th of July.	Idea paper with the title "Combining Multiple Modalities with Perceiver in Imitation-based Urban Driving"	Published
Research methods and methodology of informatics and computer engineering	Passed with score of 9/10	Planned participation at DAMSS 2021, Druskininkai, Lithuania	Planned to be performed on the 4th of December.		

Workshops participated in

Workshop	ECTS
MOKSLINIŲ REZULTATŲ PUBLIKAVIMAS PAGAL FORMALAUS VERTINIMO REIKALAVIMUS	0.1
MOKSLINĖS INFORMACIJOS IŠTEKLIAI, PAIEŠKA, IR ĮRANKIAI	0.1
MENDELEY PRAKTINIS UŽSIĖMIMAS	0.15
Total:	0.35/3

Stages of research and dissertation preparation

Vilnius University

	Name of task	Duration	Notes
1.	Review and analysis of scientific research on the topic of the dissertation (in Lithuania and abroad): 1. Defining and describing the objectives of the dissertation research topic. 2. Overview of deep imitation learning and deep reinforcement learning for mobile robot navigation. 3. Summary of methods overview and presentation on the description of the analytical part of the dissertation. 4. Formation of research goal.	September 2020 – August 2021	Performed literature review on imitation learning and navigation methods. Rest defined.
2.	Carrying out research: **2.1. Development of research methodology:** 1. Identification and specification of problems arising in currently available methods. 2. Specification of tasks to conduct which address to identified problems. 3. Specification of navigation environments which will be analysed further. 4. Selection of appropriate research methodology. 5. Planning of theoretical and empirical research.	September 2021 – October 2021	
	2.2. Theoretical research: 1. Analysis of reactive imitation learning methods for sensorimotor control and strategy functions, which utilize deep neural networks, such as behaviour cloning, inverse reinforcement learning, generative adversarial imitation learning, etc.	November 2021 – February 2022	

Research Object and Aim

Research object:

- Deep imitation learning methods.
- Application of deep imitation learning methods for mobile robot navigation.

Research aim:

- To develop, implement and research an autonomous navigation system for mobile robots based on imitation learning and deep neural networks

Objectives of Research

1. To **develop and investigate** new sensorimotor reflex algorithms based on deep neural networks and various simulation learning paradigms (e.g. behaviour cloning, generative adversarial imitation learning) (e.g. trajectory following, obstacle avoidance, approach to a recognized object).
2. To **compose and implement** a new navigation system for mobile robots from the obtained sensorimotor reflexes.
3. To **compare** the obtained navigation system with alternative robot navigation algorithms.
4. To **prepare publicly available datasets** for the research of autonomous robot navigation algorithms based on the principles of deep neural networks and imitation training.

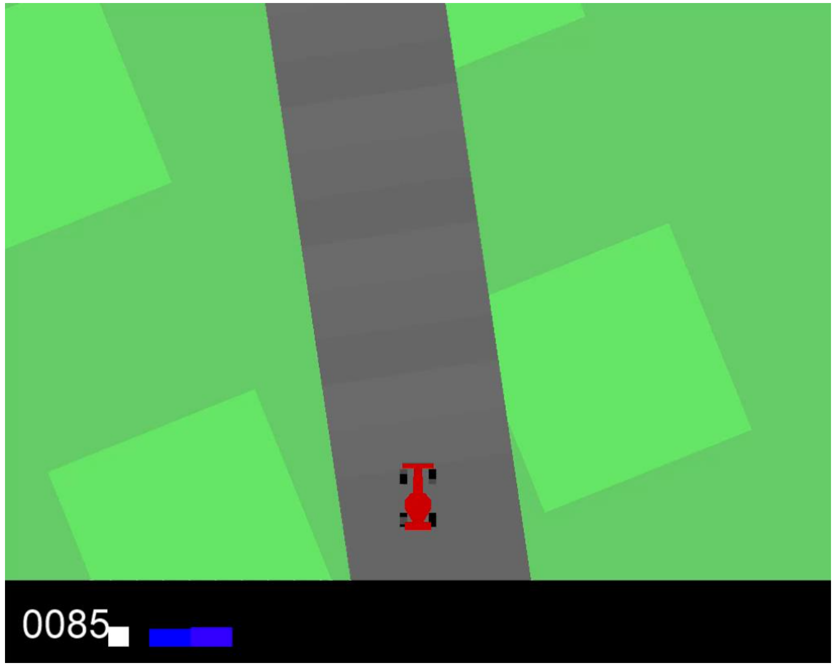
Vilnius
University

What has been carried out so far

- Literature study from papers on imitation learning for mobile robot navigation
- Took courses:
  - Machine learning (at VU)
  - Research methodology (at VU)
  - Reinforcement learning (Online)
- Trying out Simulators (CARLA and OpenAI gym)
- Attempted to run state of the art methods in simulation
- Participation in an international conference

Self generated results

Vilnius University



Literature Review

Learning to imitate

In imitation learning:

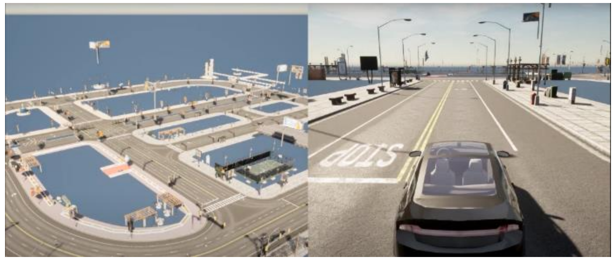
- Given: Demonstrations
- Goal: Train a policy (model) to mimic demonstrations

Being a form of machine learning, data is collected, models are optimized, accuracies are evaluated.

Vilnius University

About the problem to solve

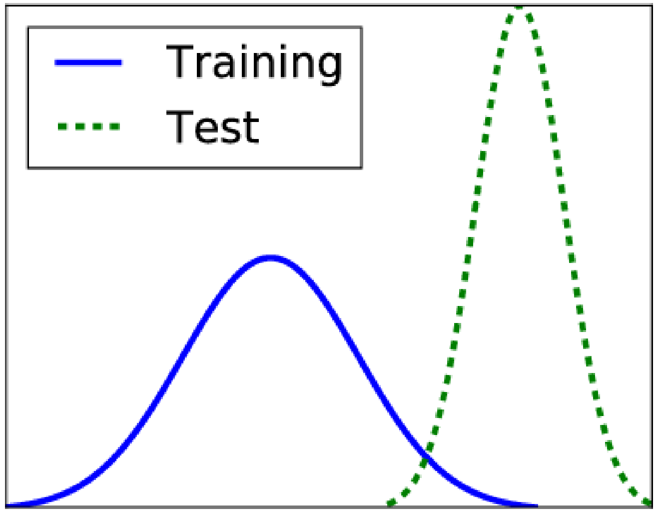
- Learning sensorimotor skills to drive and navigate based on visual input.
- It can be done with traditional methods such as SLAM, but it would require expensive sensors and extensive programming.
- The idea of imitation learning promises to solve this problem by learning from human demonstrations.
- Yet, it remains unsolved due the unpredictability of the real world causing the problem of covariate shift.
- To compare the ability between methods NoCrash benchmark has been established.
- NoCrash benchmark uses CARLA simulator to seed vehicles in different parts of a map and tests the ability of reaching from point A to B, under different sets of conditions.

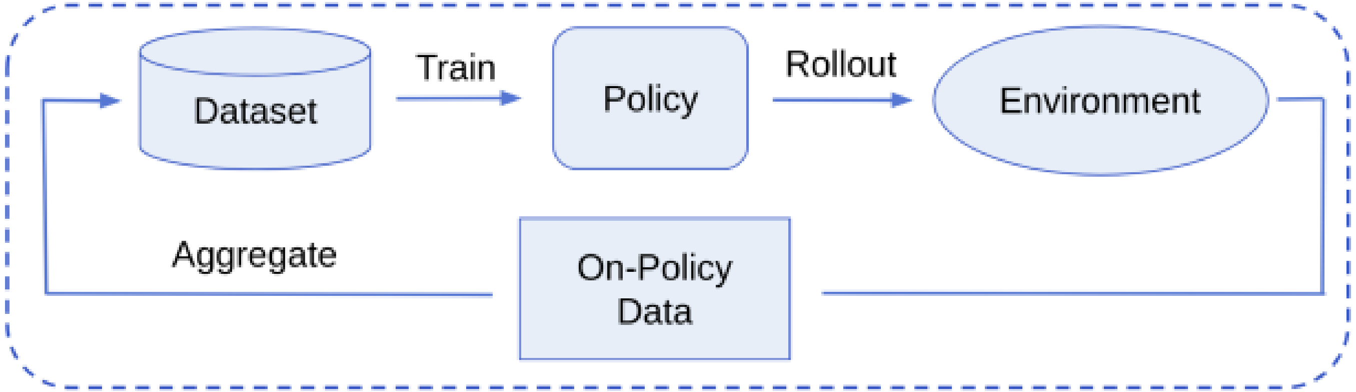


DAgger or Data Aggregation

Vilnius University

- Big issue in imitation learning is the problem of covariate shift.
- Data aggregation is a method for solving covariate shift.





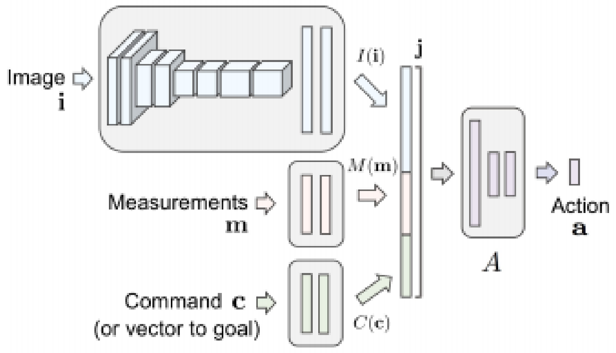
Stephane Ross, Geoffrey J. Gordon, and Drew Bagnell. A ´ reduction of imitation learning and structured prediction to no-regret online learning. In Conference on Artificial Intelligence and Statistics (AISTATS), 2011. 1, 2, 5

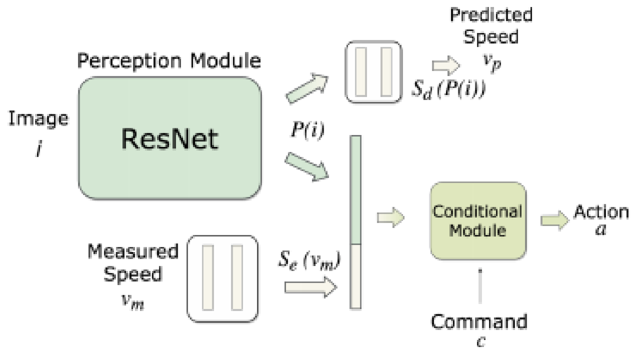
Methods for Trajectory Following (Recap)

- Conditional Imitation Learning **(CIL, 2018)** uses imitation learning with high level commands conditioned to the input to learn the skill of trajectory following.
- Conditional Affordance Learning **(CAL, 2018)** learns affordances in the form of low dimensional intermediate representations from videos, while conditioning with high level commands.
- Conditional Imitation learning with Resnet and speed branch **(CILRS, 2019)** is an extension of CIL with change in neural network architecture and using a separate branch to predict speed.
- Learning by Cheating **(LBC, 2019)** proposes training an agent in a two-step process, once with privileged information and once from a teacher network without privileged information.

Methods for Trajectory Following (Recap)

- Implicit Affordances **(IA, 2020)** uses a encoder to learn to predict affordances and then uses reinforcement learning to learn to navigate based on the affordances.
- Affordances based reinforcement learning **(IRL, 2021)** experiments with combining implicit and explicit affordances and training with reinforcement learning.





Learning situational driving

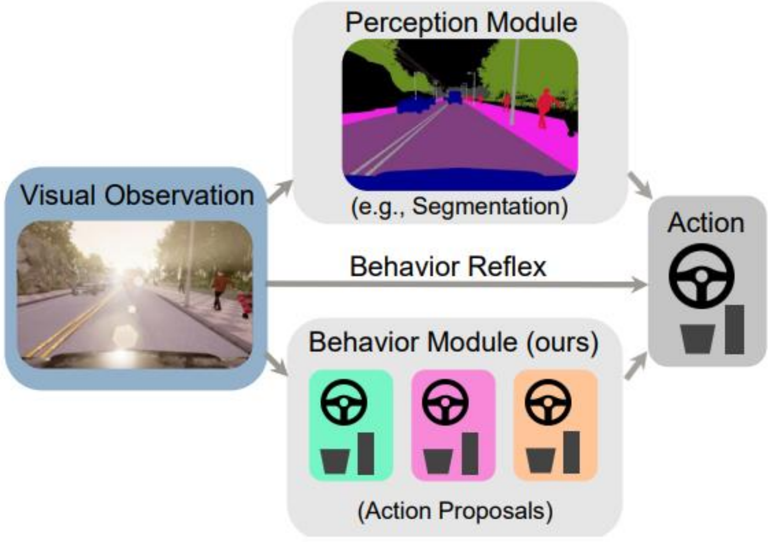
- Humans are able to drive under diverse visual conditions and situations.
- A sensorimotor model should be able to do so to achieve reliable driving.
- That is, the same model should be able to drive in several conditions:
  - A sunny day on a high way
  - A rainy day on roads where lane markings aren't visible
  - A busy intersection crowded with people
- To tackle such conditions humans are able to leverage multiple types of reasoning and learning strategies.

E. Ohn-Bar, A. Prakash, A. Behl, K. Chitta and A. Geiger, "Learning Situational Driving," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 11293-11302, doi: 10.1109/CVPR42600.2020.01131.

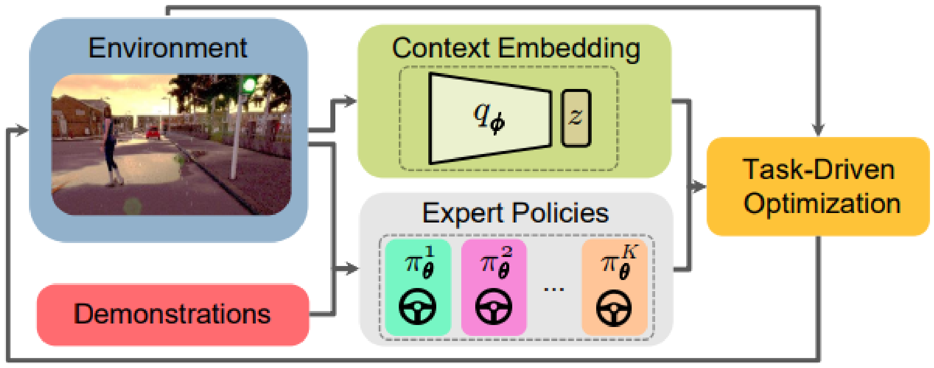
Vilnius University

Learning situational driving

- This paper proposes utilising a mixture of experts (MOE) model to model driving in various conditions.
- Along with MOE, it trains a context embedding model.
- First expert models are trained, followed by whole architecture.



$$\pi_{\Theta}(\mathbf{a}|\mathbf{o},c) = \sum_{k=1}^{K} \underbrace{\alpha_{\theta}^{k}(\mathbf{o},c)}_{\text{Mixture Weights}} \underbrace{\pi_{\theta}^{k}(\mathbf{a}|\mathbf{o},c)}_{\text{Expert Models}} + \Psi \underbrace{\begin{bmatrix} q_{\phi}(\mathbf{I}) \\ v \\ c \end{bmatrix}}_{\text{Context Embedding}}$$



DAgger with critical states and replay buffer

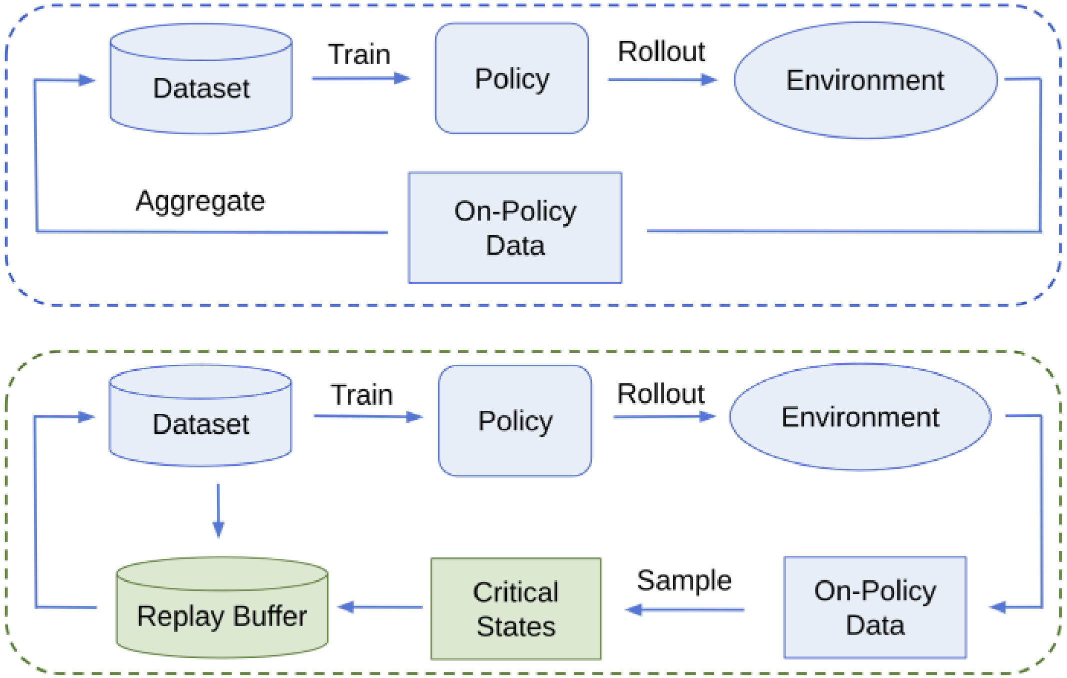
Vilnius University

- This paper challenges the methods of data aggregation and studies the current methods extensively.
- Points out how in current methods, successive iterations of data aggregation start to deteriorate the performance.
- The propose a data aggregation method which:
  - Samples critical states from collected on policy data
  - Incorporates a replay buffer which focuses on uncertainty of state distribution.

Aditya Prakash, Aseem Behl, Eshed Ohn-Bar, Kashyap Chitta, Andreas Geiger; Exploring Data Aggregation in Policy Learning for Vision-Based Urban Autonomous Driving, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 11763-11773

Vilnius University

DAgger with critical states and replay buffer



Algorithm 1 DAgger with Critical States and Replay Buffer

Collect D_0 using expert policy π^*
$\hat{\pi}_0 = \text{argmin}_\pi \mathcal{L}(\pi, \pi^*, D_0)$
Initialize replay buffer $D \leftarrow D_0$
Let $m = |D_0|$
for $i = 1$ *to* N **do**
  Generate on-policy trajectories using $\hat{\pi}_{i-1}$
  Get dataset $D_i = \{(s, \pi^*(s))\}$ of visited states by $\hat{\pi}_{i-1}$ and actions given by expert
  Get $D_i' \leftarrow \{(s_c, \pi^*(s_c))\}$ after sampling critical states from D_i
  Combine datasets: $D \leftarrow D \cup D_i'$
  while $|D| > m$ **do**
    Sample $(s, \pi^*(s))$ randomly from $D \cap D_0$
    $D \leftarrow D - \{(s, \pi^*(s))\}$
  end
  Train $\hat{\pi}_i = \text{argmin}_\pi \mathcal{L}(\pi, \pi^*, D)$ with policy initialized from $\hat{\pi}_{i-1}$
end
return $\hat{\pi}_N$

DAgger with critical states and replay buffer

Vilnius University

- **Critical states**:
  - **Task based**: Strong turns at intersections and other places, with help of CARLA.
  - **Policy based**: Uses test time dropout (value of 0.5) to evaluate which states have a high variance in prediction, sample ones above threshold.
  - **Policy and export based:**
    - States with highest loss
    - Failure cases like collisions, brakes required, etc.
- As most driving data can consist of some simple repetitive events and also some complex rare events, the **replay buffer** randomly samples states from the current dataset and replaces states with high occurrences with critical states.

Imitating a RL coach

Vilnius University

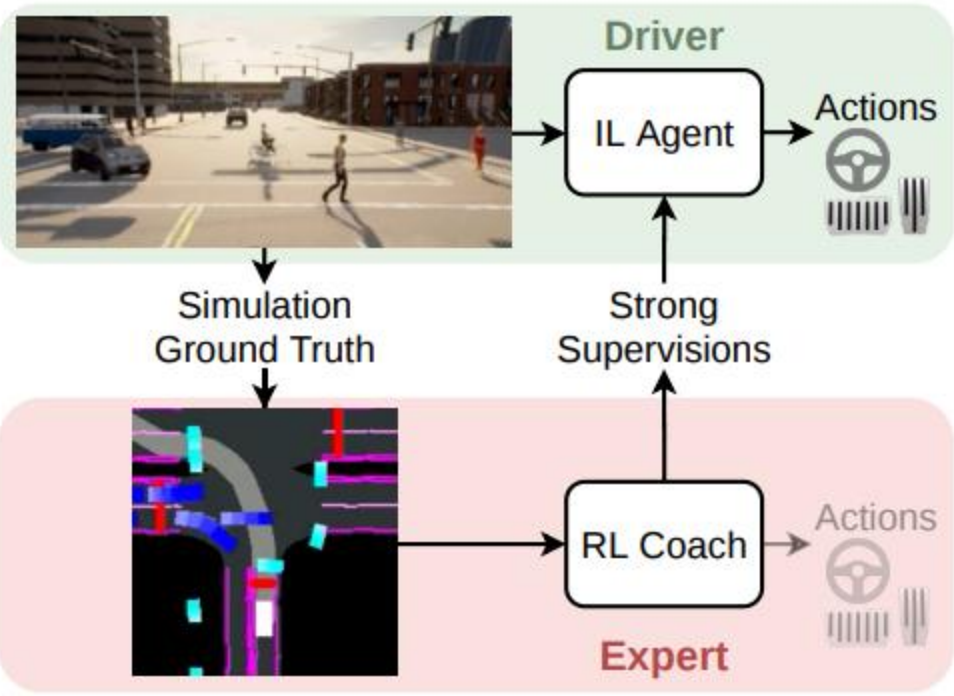
- This paper challenges the origins of dataset.
- In practice, the datasets are collected by rule based agents (autopilot) embedded in the CARLA simulator, and data from human is only collected for sparse events like interventions.
- In comparison to autopilot agents, humans drive better but learning just from humans can be inefficient.
- Therefore they break the process into a two step process by eliminating the autopilot.

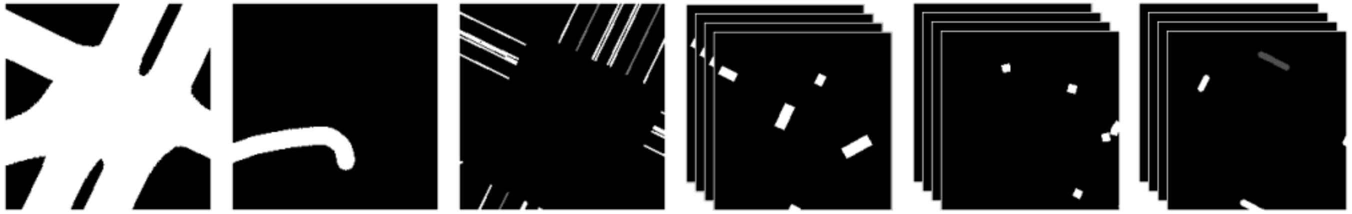
Zhang, Zhejun et al. “End-to-End Urban Driving by Imitating a Reinforcement Learning Coach.” *ArXiv* abs/2108.08265 (2021): n. pag.

Imitating a RL coach

- The problem is first modelled into a RL problem, where an agent (called the RL coach or Roach) learns to drive based on a bird’s eye view.
- Once trained, roach is then used to collect front camera data for another IL agent to be trained the traditional way.
- This way the rule based autopilot agent is eliminated.
- The paper states that the bird’s eye view provides the agent with better information and hence creates a rich dataset.

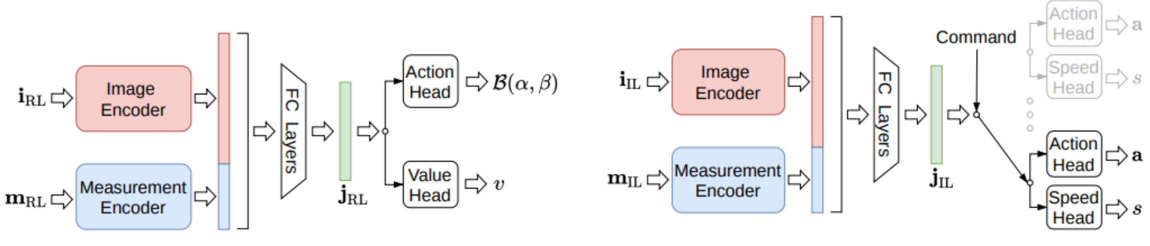
Imitating a RL coach





(a) Drivable areas (b) Desired route (c) Lane boundaries (d) Vehicles (e) Pedestrians (f) Lights and stops

Figure 2: **The BEV representation used by our Roach.**



(a) Roach (b) CILRS

Vilnius University

Published work

On record:

- Conference: All sensors 2021
- Participation type: Idea paper

Off record:

- Journal “Springer: Autonomous Robots”
- Impact factor: 3.6

Combining Multiple Modalities with Perceiver in Imitation-based Urban Driving

Shubham Juneja
Institute of Data Science &
Digital Technologies
Vilnius University
Vilnius, Lithuania
Email: shubham.juneja@mif.stud.vu.lt

Virginijus Marcinkevičius
Institute of Data Science &
Digital Technologies
Vilnius University
Vilnius, Lithuania
Email: virginijus.marcinkevicius@mif.vu.lt

Povilas Daniušis
Department of Business
Technologies & Entrepreneurship
Vilnius Gediminas Technical University
Vilnius, Lithuania
Email: povilas.daniusis@vgtu.lt

SpringerLink

Published: 04 May 2021

Topological navigation graph framework

Povilas Daniušis, Shubham Juneja, Lukas Valatka & Linas Petkevičius

Autonomous Robots (2021) | Cite this article

Vilnius University

Combining multiple modalities with Perceiver in IL based learning

- We present a study pointing out how end-to-end methods rely on a single modality while lacking the performance compared to traditional autonomous driving methods which take a modular approach.
- Therefore, we propose a method to enrol more than one modality in the learner.
- We propose the use of a perceiver architecture in the learner as this architecture shows capability of learning with varying number and types of modalities as input data.
- Since the published paper is a idea paper, no experiments were presented.

Work plan for semester 3

Vilnius University

Review and analysis of scientific research on the topic of the dissertation (in Lithuania and abroad):

- Development of Research methodology
- Initiating analytical research

Passing exam:

- Fundamentals of informatics and informatics engineering
- Optimization methods and their applications

Publication plan:

- Review of research on topic of the dissertation (in conference proceedings)

Conference Participation:

- Participation in DAMSS

Vilnius
University

Thank you