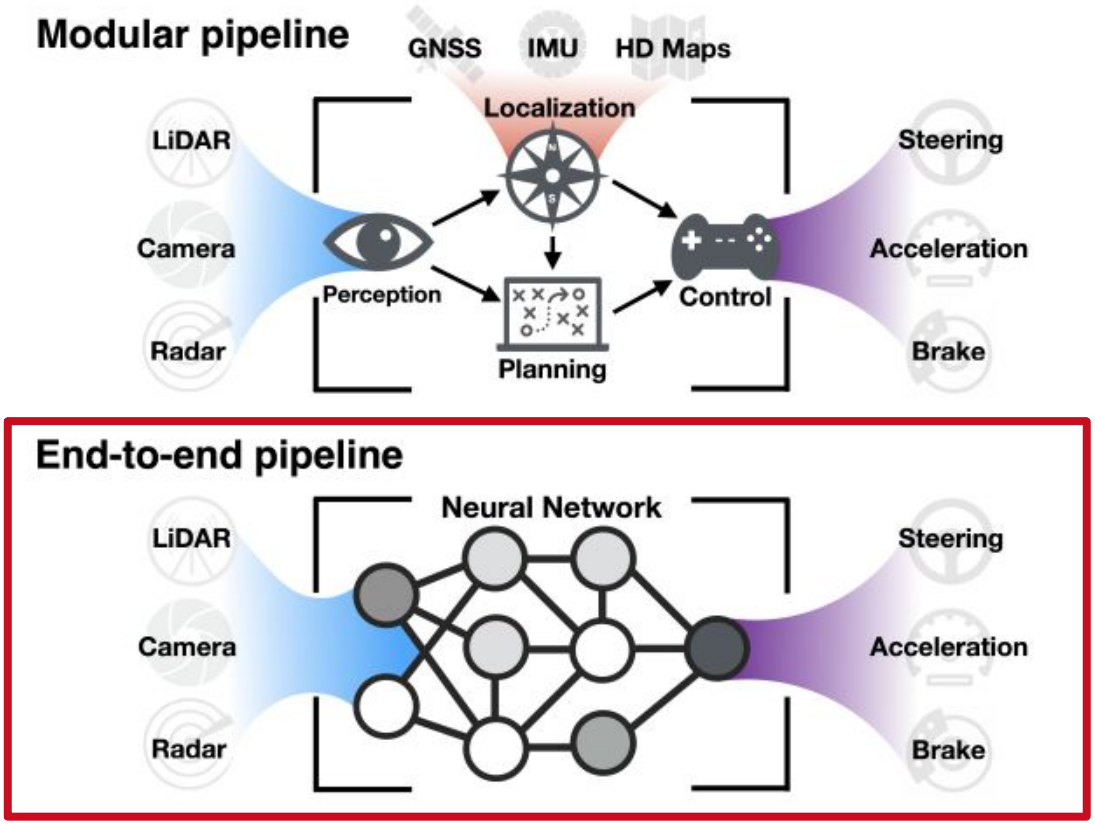
End to End Multimodal Imitation Learning

Group 3 | APP-RAS | Milestone 3 | WS 22/23

Agenda

1. **End-to-End Imitation Learning in Autonomous Driving**
2. **Goal & Steps Overview**
3. **Dataset**
4. **Model Architectures**
5. **Model Training**
6. **Model Evaluation**

1. End-to-End Multimodal Imitation Learning in AD

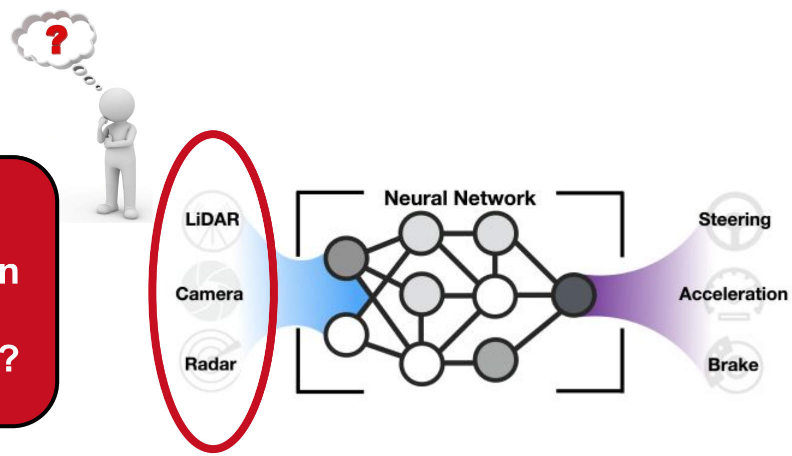


https://pantelis.github.io/cs677/docs/common/lectures/autonomous-cars/

Technische Universität Berlin

2. Goal & Steps Overview

How helpful are different sensors in End-to-End Imitation Learning?



Goal: Try different combination!

Data Acquisition → **Data Preprocessing** → **Architecture Selection** → **Model Training** → **Model Evaluation**

https://pantelis.github.io/cs677/docs/common/lectures/autonomous-cars/

Technische Universität Berlin

3. Dataset - Acquisition

Using CARLA Simulator for Data Acquisition

Dataset

downloaded

95,6%

Dataset readily available from "TransFuser" Paper

generated

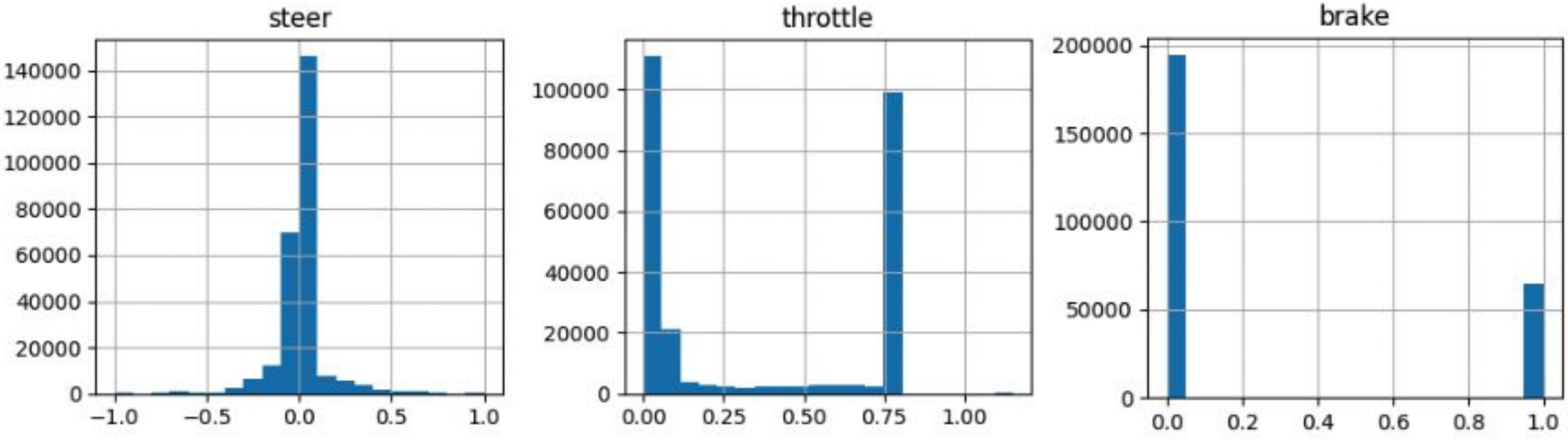
4,4%

Additional Noise-Injected Data using "Expert Driver" of "TransFuser" Paper

Chitta et al.; TransFuser: Imitation with Transformer-Based Sensor Fusion for Autonomous Driving (2022)

3. Dataset - Characteristics

- Driving Length: 1 day 14h
- Raw size: 61 GB
- Number Towns: 9
- Sensors: 3 RGB (160 x 960), LiDAR, Speedometer, …
- Target values: Steer, Throttle, Brake (Control Commands)
- Additional: Navigational Command (turn left, turn right, go straight, …)



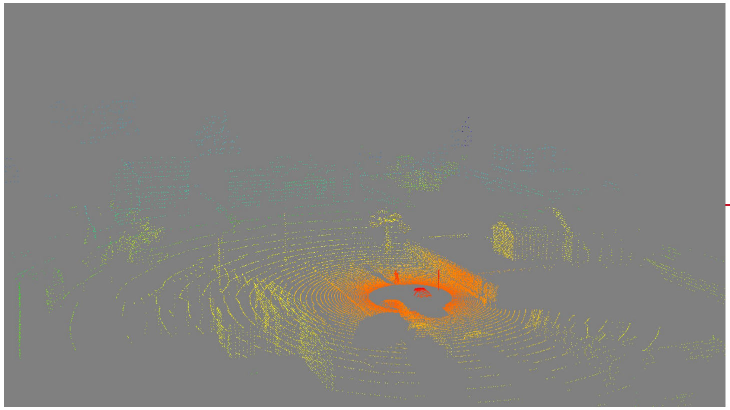
Technische Universität Berlin

3. Dataset - Preprocessing

RGB

Resize to 88 x 224 x 3 → Scale to [0, 1] → Z - Normalization

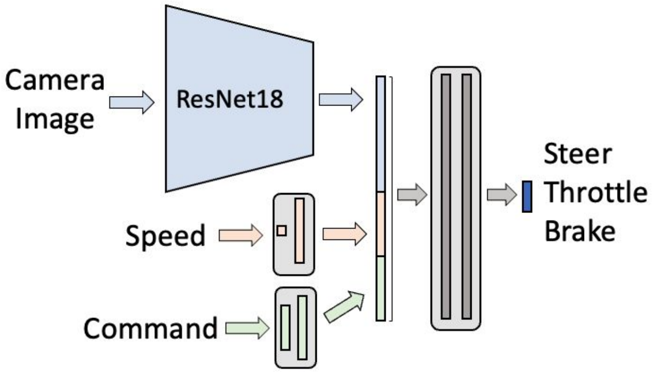
LiDAR



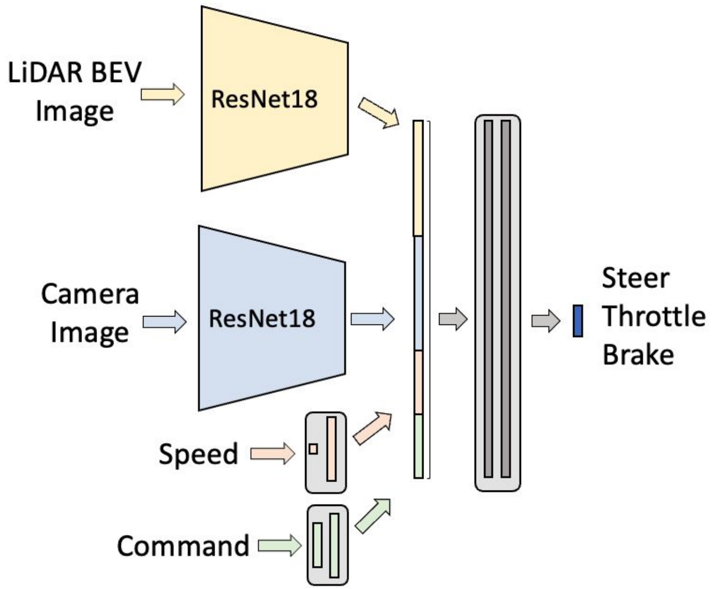
Scale to [0, 1] → Z - Normalization

4. Model Architectures

Camera



Camera + LiDAR



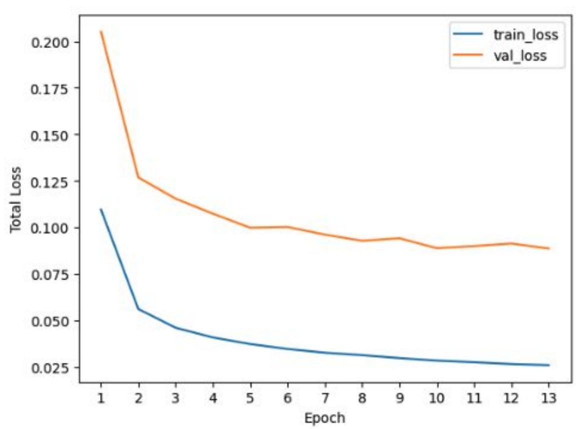
Codevilla et al.; End-to-end Driving via Conditional Imitation Learning (2017)

5. Model Training

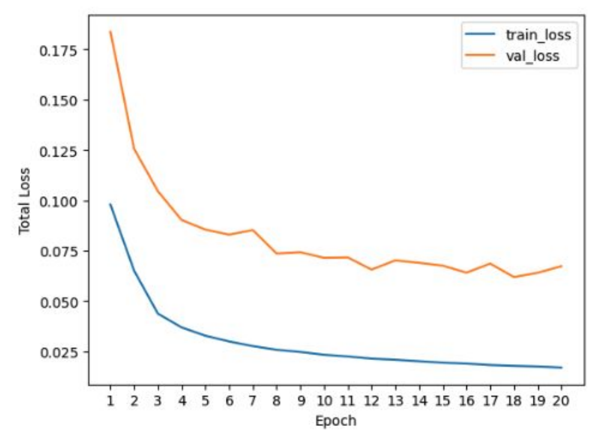
Settings

- **Fixed**
  - Adam Optimizer
  - lr = 0.0001
  - Loss functions (L1) & weights
- **Experimental**
  - Dropout
  - Data Augmentation
  - Data Balancing

Camera



Camera + LiDAR



Technische Universität Berlin

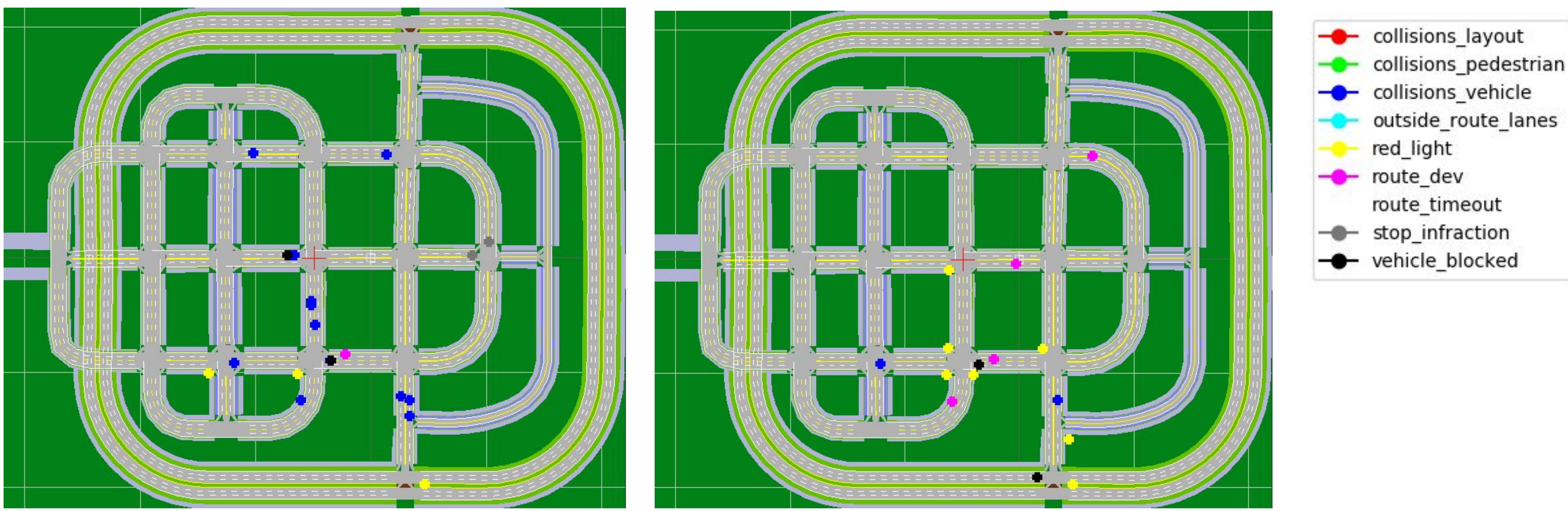
6. Model Evaluation - “Longest6” Benchmark Statistics

	Camera	Camera + LiDAR
Avg. driving score ↑	9.601	**13.365**
Avg. route completion ↑	**39.452**	33.597
Avg. infraction penalty ↑	0.358	**0.464**
Collisions with pedestrians ↓	0.000	0.000
Collisions with vehicles ↓	3.031	**1.601**
Collisions with layout ↓	0.421	**0.378**
Red lights infractions ↓	**0.512**	0.608
Stop sign infractions ↓	0.162	**0.123**
Off-road infractions ↓	0.855	**0.273**
Route deviations ↓	**0.389**	0.828
Route timeouts ↓	0.097	**0.032**
Agent blocked ↓	**0.943**	0.956

→ Camera + LiDAR outperforms Camera in the “Average driving Score”

6. Model Evaluation - Visual Inspection

Camera | **Camera + LiDAR**



6. Model Evaluation - Camera + Lidar during Night

6. Model Evaluation - Camera + Lidar - Right Turn

6. Model Evaluation - Camera - Alignment to the Lane

6. Model Evaluation - Camera - Crash during Right Turn

6. Model Evaluation - Camera - Crash in Straight Lane

Conclusion

Objectives

- **Sensor choices**: Decide on a set of sensors and different combinations (setups).
- **Expert Driver**: Choose an expert driver for CARLA.
- **Dataset**: Generate a Dataset with the expert driver.
- **E2E Model**: Implement an E2E model capable of driving autonomously.
- **Data Preprocessing** (for all sensor setups): Build a pipeline to generate compatible data for the chosen architecture.
- **Training**: Train the different models with varying sensor setups and varying data sizes.
- **Evaluation**: Conduct the experiments and compare the sensor setups.

Use Cases

We expect **route completions and low infractions** in:

- Easy environment: (short routes, less cars) empty roads, good weather (on training routes)
- Hard environment: (long routes, max cars) “Longest6”

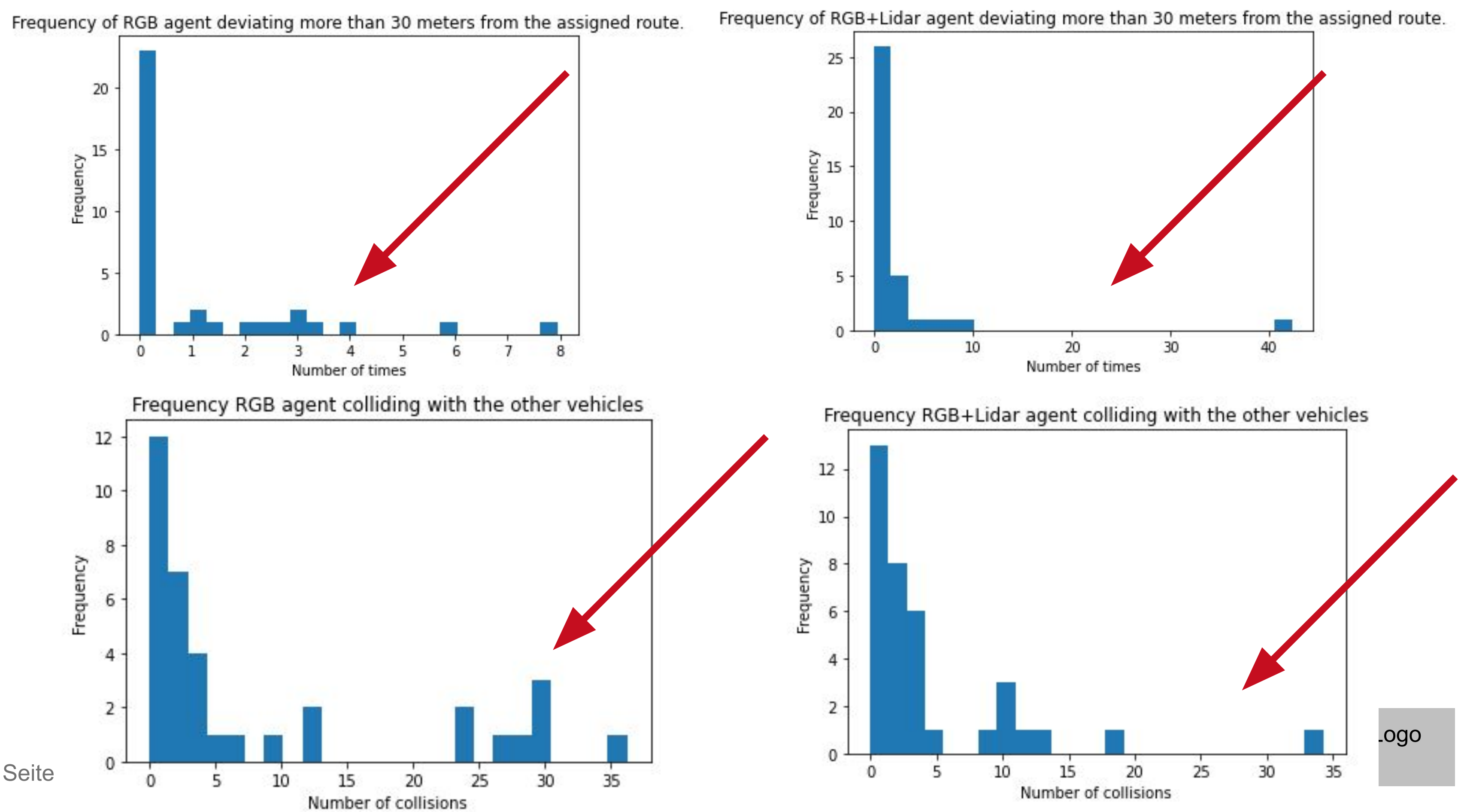
Future Work

- Room for improvement in terms of hyperparameters (functions architectures)
- Validation using CARLA after every epoch (Model Selection)
- Changing to Vision Transformer architectures
- More sensor configurations
- More in-depth comparative analysis and more context

Thank you for listening!

Any questions?

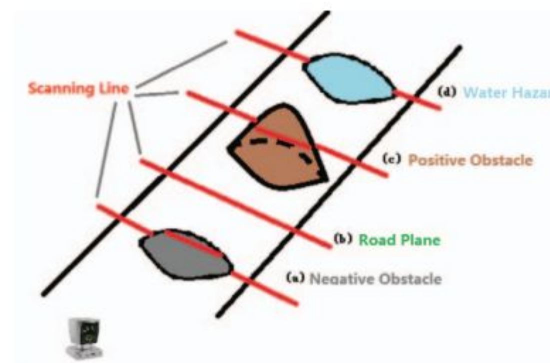
Visualising Dev and Collision improvement



Logo

Lidar Preprocessing

- Transform [X ,Y, Z , I] (coordinates, intensity loss during travel)
- To 2d Histogram representing a horizontal plane in front of the car
- Birds eye view, can process both other cars in traffic (positive obstacles), as well as holes in the ground or small objects below the eye level (negative obstacles)
- “Lidar-histogram for fast road and obstacle detection”



Leaderboard

How do we measure performance, specifically?

We use the Longest6 benchmark!

- 36 routes
- Average route length 1.5 km (official leaderboard ~1.7km)
- High density of dynamic agents: vehicles spawned at all possible locations in CARLA
- Unique environmental conditions with combinations of:
  - Six weather conditions: Cloudy, Wet, MidRain, WetCloudy, HardRain, SoftRain
  - Six daylight conditions Night, Twilight, Dawn, Morning, Noon, Sunset.

Dezentrales Logo optional

Technische Universität Berlin

Measuring performance

The Longest6 benchmark uses three major factors to measure performance:

- **Driving score**: RiPi, — Main metric of the leaderboard, serving as the product between the route completion and the infractions penalty. Here **Ri** is the percentage of completion of the i−th route, and **Pi**, the infraction penalty of the i−th route.
- **Route completion**: Percentage of the route distance completed by an agent.
- **Infraction penalty**: $\prod_{j}^{ped., ..., stop}(p_i^j)^{\#infractions}$.The leaderboard tracks several types of infractions and this metric aggregates all of these infractions triggered by an agent as a geometric series. Agents start with an ideal **1.0** base score, which is reduced each type an infraction is commited.

Dezentrales Logo optional

Breaking it down

Infractions and shutdown events

The CARLA leaderboard offers individual metrics for a series of infractions. Each of these has a penalty coefficient that will be applied everytime it happens. Ordered by severity, the infractions are the following.

- **Collisions with pedestrians** — 0.50.
- **Collisions with other vehicles** — 0.60.
- **Collisions with static elements** — 0.65.
- **Running a red light** — 0.70.
- **Running a stop sign** — 0.80.

Dezentrales Logo optional