KE:SAI

Open Science Autonomy Lab

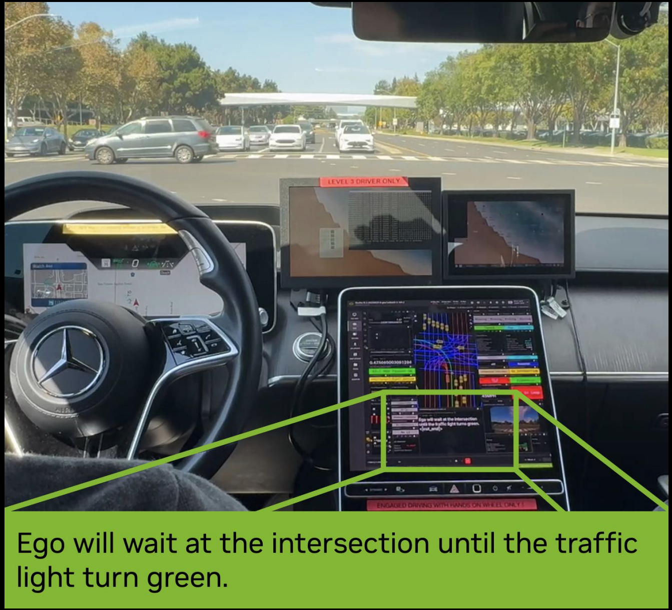
Democratizing Autonomous Driving with **Open Science**

kashyap7x.github.io

Our mission is to advance the efficiency frontier of robust and safe physical AI through fully open and reproducible research.

Current frontier: Vision Language Action models

Predicting an **expert action** conditioned on a **visual world state** and a **reasoning trace** in **natural language**



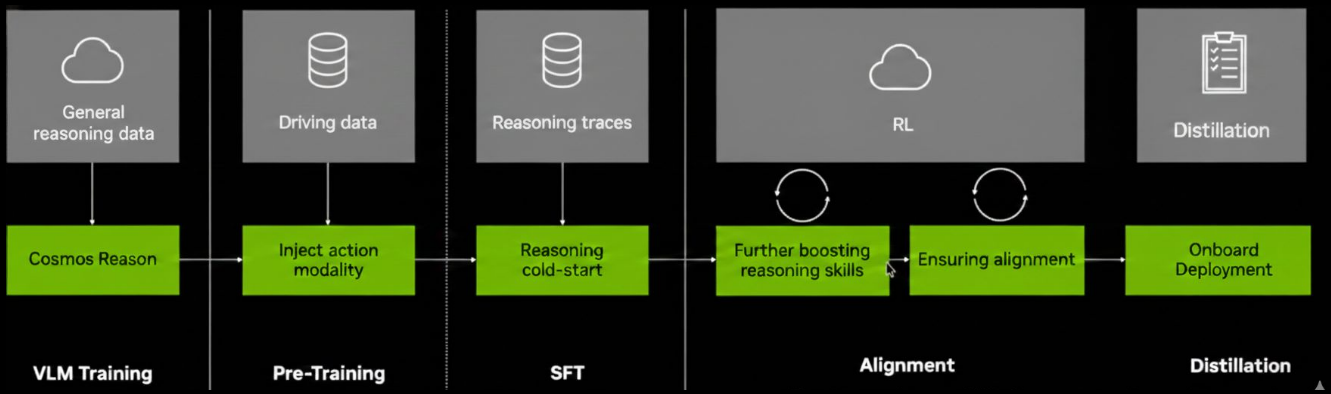
- **Reasoning, rare case handling, arbitration**
- Semi-automatic labeling
- **Misalignment** of reasoning to actions
- Hard to incorporate **full sensor suite**
- Many parameters dedicated to **knowledge memorization**
- …

Wang et al., Alpamayo-R1: Bridging Reasoning and Action Prediction for Generalizable Autonomous Driving in the Long Tail. ArXiv 2025

The current frontier is inefficient

Vision Language Action models are the dominant paradigm since 2025, but they are

- Data inefficient (80k hours of curated expert data)
- Compute inefficient (e.g., 60k GPU hours for in-domain “Pre-Training”)
- Inference inefficient (100ms on server GPU after quantization and distillation)



Wang et al., Alpamayo-R1: Bridging Reasoning and Action Prediction for Generalizable Autonomous Driving in the Long Tail. ArXiv 2025

Open
Datasets

We need heterogeneity & diversity

Training LLMs

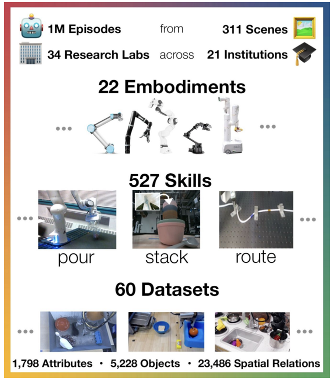
COMMON CRAWL

Datasets

huggingface.co/datasets

... or generalist robots

LeRobot




Physical AI AV Dataset


OpenDV-2k

Open
Tools

Driving datasets in a nutshell

Download and setup various file formats.

```
$AV2_DATA_ROOT
└── sensor/
    ├── train/
    │   └── 00a6ffc1-6ce9-3bc3-a060-6006e9893a1a/
    │       ├── annotations.feather
    │       ├── calibration/
    │       │   ├── egovehicle_SE3_sensor.feather
    │       │   └── intrinsics.feather
    │       ├── city_SE3_egovehicle.feather
    │       ├── map/
    │       │   └── ...
    │       └── sensors/
    │           ├── cameras/...
    │           └── lidar/...
    ├── val/
    └── test/
```

```
$NCORE_DATA_ROOT
└── clips/
    └── {clip_id}/
        ├── pai_{clip_id}.json
        ├── pai_{clip_id}.ncore4.zarr.itar
        ├── pai_{clip_id}.ncore4-lidar_top_360fov.zarr.itar
        └── pai_{clip_id}.ncore4-camera_{name}.zarr.itar
```

```
$PANDASET_DATA_ROOT/
├── 001/
│   ├── annotations/
│   │   ├── cuboids/
│   │   │   ├── 00.pkl.gz
│   │   │   ├── ...
│   │   │   └── 79.pkl.gz
│   │   └── semseg/  (currently not used)
│   │       ├── 00.pkl.gz
│   │       ├── ...
│   │       ├── 79.pkl.gz
│   │       └── classes.json
│   ├── camera/
│   │   ├── back_camera/
│   │   │   ├── 00.jpg
│   │   │   ├── ...
│   │   │   ├── 79.jpg
│   │   │   ├── intrinsics.json
│   │   │   ├── poses.json
│   │   │   └── timestamps.json
│   │   ├── front_camera/
│   │   │   └── ...
│   │   ├── front_left_camera/
│   │   │   └── ...
│   │   ├── front_right_camera/
│   │   │   └── ...
│   │   ├── left_camera/
│   │   │   └── ...
│   │   └── right_camera/
│   │       └── ...
│   ├── LICENSE.txt
│   ├── lidar/
│   │   ├── 00.pkl.gz
│   │   ├── ...
│   │   ├── 79.pkl.gz
│   │   ├── poses.json
│   │   └── timestamps.json
│   └── meta/
│       ├── gps.json
│       └── timestamps.json
├── ...
└── 158/
    └── ...
```

```
$NUSCENES_DATA_ROOT
├── can_bus/
│   ├── scene-0001_meta.json
│   ├── ...
│   └── scene-1110_zoe_veh_info.json
├── maps/
│   ├── 36092f0b03a857c6a3403e25b4b7aab3.png
│   ├── ...
│   ├── 93406b464a165eaba6d9de76ca09f5da.png
│   ├── basemap/
│   │   └── ...
│   ├── expansion/
│   │   └── ...
│   └── prediction/
│       └── ...
├── samples/
│   ├── CAM_BACK/
│   │   └── ...
│   ├── ...
│   └── RADAR_FRONT_RIGHT/
│       └── ...
├── sweeps/
│   └── ...
├── v1.0-mini/
│   ├── attribute.json
│   ├── ...
│   └── visibility.json
├── v1.0-test/
│   ├── attribute.json
│   ├── ...
│   └── visibility.json
└── v1.0-trainval/
    ├── attribute.json
    ├── ...
    └── visibility.json
```

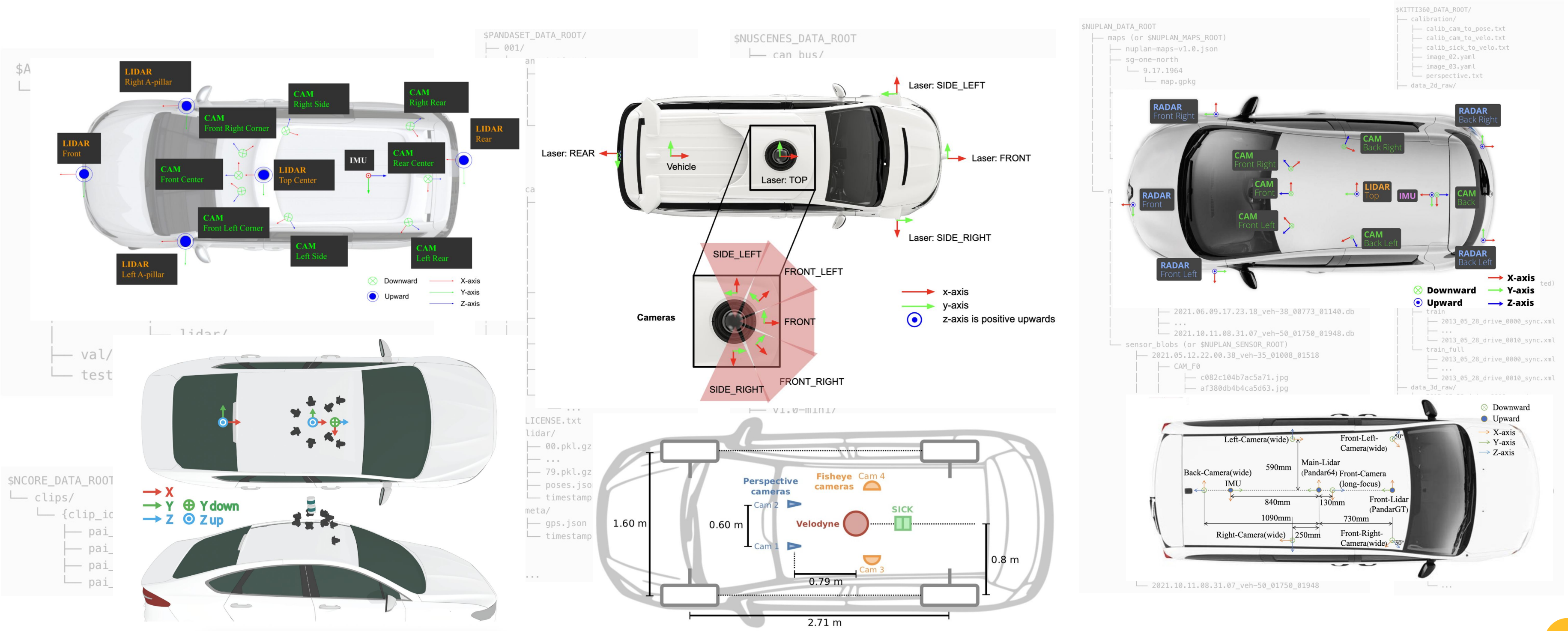
```
$NUPLAN_DATA_ROOT
├── maps (or $NUPLAN_MAPS_ROOT)
│   ├── nuplan-maps-v1.0.json
│   ├── sg-one-north
│   │   └── 9.17.1964
│   │       └── map.gpkg
│   ├── us-ma-boston
│   │   └── 9.12.1817
│   │       └── map.gpkg
│   ├── us-nv-las-vegas-strip
│   │   └── 9.15.1915
│   │       └── map.gpkg
│   └── us-pa-pittsburgh-hazelwood
│       └── 9.17.1937
│           └── map.gpkg
└── nuplan-v1.1
    ├── splits
    │   ├── mini
    │   │   ├── 2021.05.12.22.00.38_veh-35_01008_01518.db
    │   │   ├── 2021.06.09.17.23.18_veh-38_00773_01140.db
    │   │   ├── ...
    │   │   └── 2021.10.11.08.31.07_veh-50_01750_01948.db
    │   ├── test
    │   │   └── ...
    │   └── trainval
    │       ├── 2021.05.12.22.00.38_veh-35_01008_01518.db
    │       ├── 2021.06.09.17.23.18_veh-38_00773_01140.db
    │       ├── ...
    │       └── 2021.10.11.08.31.07_veh-50_01750_01948.db
    └── sensor_blobs (or $NUPLAN_SENSOR_ROOT)
        ├── 2021.05.12.22.00.38_veh-35_01008_01518
        │   ├── CAM_F0
        │   │   ├── c082c104b7ac5a71.jpg
        │   │   ├── af380db4b4ca5d63.jpg
        │   │   ├── ...
        │   │   └── 2270fccfb44858b3.jpg
        │   ├── CAM_B0
        │   ├── CAM_L0
        │   ├── CAM_L1
        │   ├── CAM_L2
        │   ├── CAM_R0
        │   ├── CAM_R1
        │   ├── CAM_R2
        │   └──MergedPointCloud
        │       ├── 03fafcf2c0865668.pcd
        │       ├── 5aee37ce29665f1b.pcd
        │       ├── ...
        │       └── 5fe65ef6a97f5caf.pcd
        │
        ├── 2021.06.09.17.23.18_veh-38_00773_01140
        ├── ...
        └── 2021.10.11.08.31.07_veh-50_01750_01948
```

```
$KITTI360_DATA_ROOT/
├── calibration/
│   ├── calib_cam_to_pose.txt
│   ├── calib_cam_to_velo.txt
│   ├── calib_sick_to_velo.txt
│   ├── image_02.yaml
│   ├── image_03.yaml
│   └── perspective.txt
├── data_2d_raw/
│   ├── 2013_05_28_drive_0000_sync/
│   │   ├── image_00/
│   │   │   ├── data_rect
│   │   │   │   ├── 0000000000.png
│   │   │   │   ├── ...
│   │   │   │   └── 0000011517.png
│   │   │   └── timestamps.txt
│   │   ├── image_01/
│   │   │   └── ...
│   │   ├── image_02/
│   │   │   ├── data_rgb
│   │   │   │   ├── 0000000000.png
│   │   │   │   ├── ...
│   │   │   │   └── 0000011517.png
│   │   │   └── timestamps.txt
│   │   └── image_03/
│   │       └── ...
│   ├── ...
│   └── 2013_05_28_drive_0010_sync/
│       └── ...
├── data_2d_semantics/ (not yet supported)
│   └── ...
├── data_3d_bboxes/
│   ├── train
│   │   ├── 2013_05_28_drive_0000_sync.xml
│   │   ├── ...
│   │   └── 2013_05_28_drive_0010_sync.xml
│   └── train_full
│       ├── 2013_05_28_drive_0000_sync.xml
│       ├── ...
│       └── 2013_05_28_drive_0010_sync.xml
├── data_3d_raw/
│   ├── 2013_05_28_drive_0010_sync/
│   │   └── velodyne_points/
│   │       ├── data
│   │       │   ├── 0000000000.bin
│   │       │   ├── ...
│   │       │   └── 0000011517.bin
│   │       └── timestamps.txt
│   ├── ...
│   └── 2013_05_28_drive_0010_sync/
│       └── ...
├── data_3d_semantics/ (not yet supported)
│   └── ...
└── data_poses/
    ├── 2013_05_28_drive_0000_sync/
    │   ├── cam0_to_world.txt
    │   ├── oxts/
    │   │   └── ...
    │   └── poses.txt
    ├── ...
    └── 2013_05_28_drive_0010_sync/
        └── ...
```


Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

Driving datasets in a nutshell

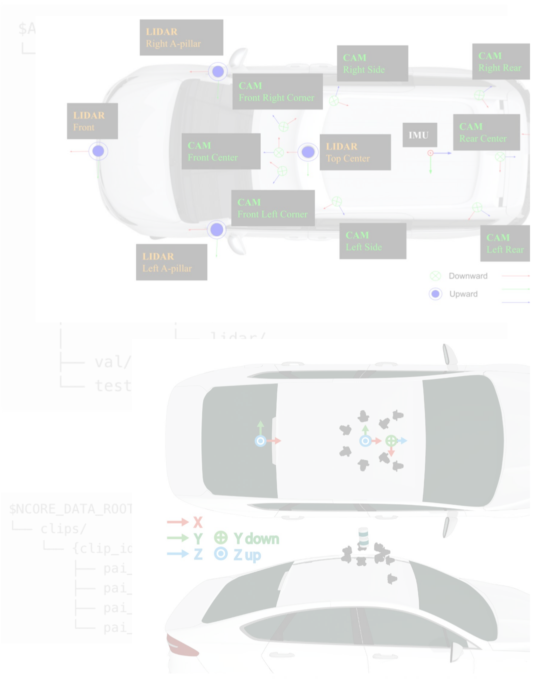
Implement and handle multiple sensors and coordinate conventions



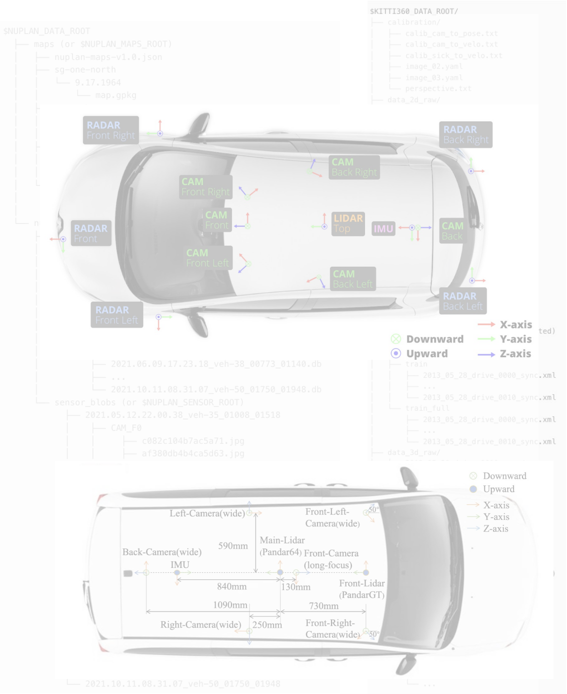
Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

Driving datasets in a nutshell

Adapt various semantic classes for annotations.



Dataset	#	Original Labels (grouped by default category)
nuScenes [6]	23	*Vehicle:* vehicle.car, vehicle.truck, vehicle.bus.bendy, vehicle.bus.rigid, vehicle.construction, vehicle.emergency.ambulance, vehicle.emergency.police, vehicle.trailer. *Two-Wheeler:* vehicle.bicycle, vehicle.motorcycle. *Person:* human.pedestrian.adult, human.pedestrian.child, human.pedestrian.construction_worker, human.pedestrian.personal_mobility, human.pedestrian.police_officer, human.pedestrian.stroller, human.pedestrian.wheelchair. *Animal:* animal. *Traffic Cone:* movable_object.trafficcone. *Barrier:* movable_object.barrier. *Generic Obj.:* movable_object.pushable_pullable, movable_object.debris, static_object.bicycle_rack.
WOD-Perc. [60]	5	*Vehicle:* type_vehicle. *Two-Wheeler:* type_cyclist. *Person:* type_pedestrian. *Traffic Sign:* type_sign. *Other:* type_unknown.
AV2-Sens. [70]	30	*Vehicle:* articulated_bus, box_truck, bus, large_vehicle, message_board_trailer, regular_vehicle, school_bus, traffic_light_trailer, truck, truck_cab, vehicular_trailer. *Train:* railed_vehicle. *Two-Wheeler:* bicycle, motorcycle. *Person:* bicyclist, motorcyclist, official_signaler, pedestrian, wheeled_rider. *Animal:* animal, dog. *Traffic Sign:* mobile_pedestrian_crossing_sign, sign, stop_sign. *Traffic Cone:* construction_barrel, construction_cone. *Barrier:* bollard. *Other:* stroller, wheelchair, wheeled_device.
PandaSet [71]	27	*Vehicle:* bus, car, emergency_vehicle, medium_sized_truck, other_vehicle_construction_vehicle, other_vehicle_pedicab, other_vehicle_uncommon, pickup_truck, semi_truck, towed_object. *Train:* train, tram_subway. *Two-Wheeler:* bicycle, motorcycle, motorized_scooter. *Person:* pedestrian, pedestrian_with_object. *Animal:* animals_bird, animals_other. *Traffic Sign:* construction_signs, signs. *Traffic Cone:* cones, pylons. *Barrier:* road_barriers, temporary_construction_barriers. *Generic Obj.:* rolling_containers. *Other:* personal_mobility_device.
KITTI-360 [48]	19	*Vehicle:* bus, car, caravan, trailer, truck. *Train:* train. *Two-Wheeler:* bicycle, motorcycle. *Person:* person, rider. *Traffic Sign:* stop, traffic_sign. *Traffic Light:* traffic_light. *Generic Obj.:* box, lamp, pole, smallpole, trash_bin, vending_machine.
WOD-Mot. [23]	5	*Vehicle:* type_vehicle. *Two-Wheeler:* type_cyclist. *Person:* type_pedestrian. *Other:* type_unset, type_other.
nuPlan [37]	7	*Vehicle:* vehicle. *Two-Wheeler:* bicycle. *Person:* pedestrian. *Traffic Sign:* czone_sign. *Traffic Cone:* traffic_cone. *Barrier:* barrier. *Generic Obj.:* generic_object.
PAI-AV [53]	12	*Vehicle:* automobile, bus, heavy_truck, other_vehicle, trailer, trolley_bus. *Train:* train_or_tram_car. *Two-Wheeler:* rider. *Person:* person, stroller. *Animal:* animal. *Generic Obj.:* protruding_object.

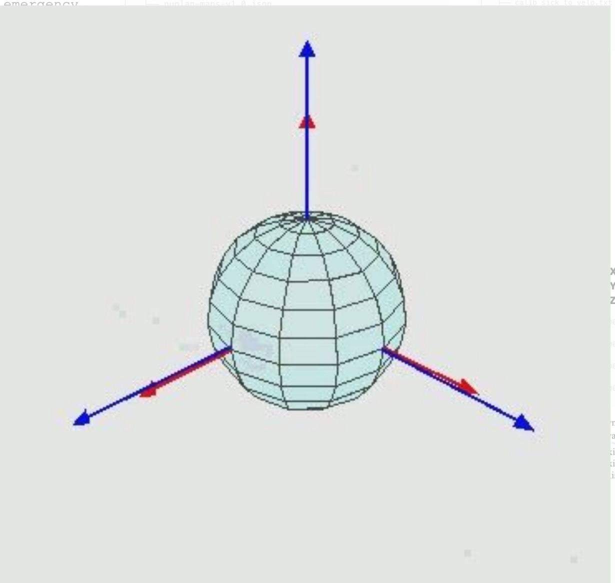


Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

Driving datasets in a nutshell

Learn about 12-24 different euler angle rotations and coordinate bugs.

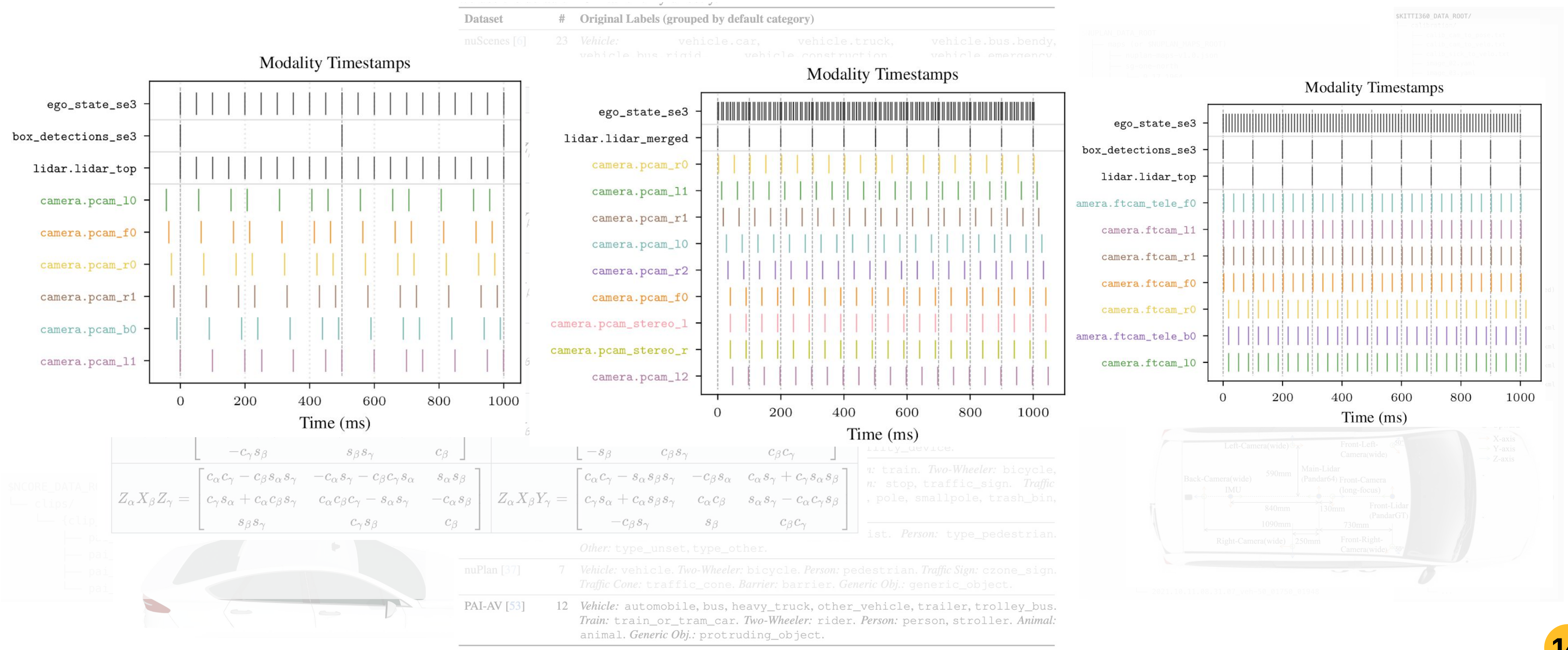
Proper Euler angles	Tait-Bryan angles
$X_\alpha Z_\beta X_\gamma = \begin{bmatrix} c_\beta & -c_\gamma s_\beta & s_\beta s_\gamma \\ c_\alpha s_\beta & c_\alpha c_\beta c_\gamma - s_\alpha s_\gamma & -c_\gamma s_\alpha - c_\alpha c_\beta s_\gamma \\ s_\alpha s_\beta & c_\alpha s_\gamma + c_\beta c_\gamma s_\alpha & c_\alpha c_\gamma - c_\beta s_\alpha s_\gamma \end{bmatrix}$	$X_\alpha Z_\beta Y_\gamma = \begin{bmatrix} c_\beta c_\gamma & -s_\beta & c_\beta s_\gamma \\ s_\alpha s_\gamma + c_\alpha c_\gamma s_\beta & c_\alpha c_\beta & c_\alpha s_\beta s_\gamma - c_\gamma s_\alpha \\ c_\gamma s_\alpha s_\beta - c_\alpha s_\gamma & c_\beta s_\alpha & c_\alpha c_\gamma + s_\alpha s_\beta s_\gamma \end{bmatrix}$
$X_\alpha Y_\beta X_\gamma = \begin{bmatrix} c_\beta & s_\beta s_\gamma & c_\gamma s_\beta \\ s_\alpha s_\beta & c_\alpha c_\gamma - c_\beta s_\alpha s_\gamma & -c_\alpha s_\gamma - c_\beta c_\gamma s_\alpha \\ -c_\alpha s_\beta & c_\gamma s_\alpha + c_\alpha c_\beta s_\gamma & c_\alpha c_\beta c_\gamma - s_\alpha s_\gamma \end{bmatrix}$	$X_\alpha Y_\beta Z_\gamma = \begin{bmatrix} c_\beta c_\gamma & -c_\beta s_\gamma & s_\beta \\ c_\alpha s_\gamma + c_\gamma s_\alpha s_\beta & c_\alpha c_\gamma - s_\alpha s_\beta s_\gamma & -c_\beta s_\alpha \\ s_\alpha s_\gamma - c_\alpha c_\gamma s_\beta & c_\gamma s_\alpha + c_\alpha s_\beta s_\gamma & c_\alpha c_\beta \end{bmatrix}$
$Y_\alpha X_\beta Y_\gamma = \begin{bmatrix} c_\alpha c_\gamma - c_\beta s_\alpha s_\gamma & s_\alpha s_\beta & c_\alpha s_\gamma + c_\beta c_\gamma s_\alpha \\ s_\beta s_\gamma & c_\beta & -c_\gamma s_\beta \\ -c_\gamma s_\alpha - c_\alpha c_\beta s_\gamma & c_\alpha s_\beta & c_\alpha c_\beta c_\gamma - s_\alpha s_\gamma \end{bmatrix}$	$Y_\alpha X_\beta Z_\gamma = \begin{bmatrix} c_\alpha c_\gamma + s_\alpha s_\beta s_\gamma & c_\gamma s_\alpha s_\beta - c_\alpha s_\gamma & c_\beta s_\alpha \\ c_\beta s_\gamma & c_\beta c_\gamma & -s_\beta \\ c_\alpha s_\beta s_\gamma - c_\gamma s_\alpha & c_\alpha c_\gamma s_\beta + s_\alpha s_\gamma & c_\alpha c_\beta \end{bmatrix}$
$Y_\alpha Z_\beta Y_\gamma = \begin{bmatrix} c_\alpha c_\beta c_\gamma - s_\alpha s_\gamma & -c_\alpha s_\beta & c_\gamma s_\alpha + c_\alpha c_\beta s_\gamma \\ c_\gamma s_\beta & c_\beta & s_\beta s_\gamma \\ -c_\alpha s_\gamma - c_\beta c_\gamma s_\alpha & s_\alpha s_\beta & c_\alpha c_\gamma - c_\beta s_\alpha s_\gamma \end{bmatrix}$	$Y_\alpha Z_\beta X_\gamma = \begin{bmatrix} c_\alpha c_\beta & s_\alpha s_\gamma - c_\alpha c_\gamma s_\beta & c_\gamma s_\alpha + c_\alpha s_\beta s_\gamma \\ s_\beta & c_\beta c_\gamma & -c_\beta s_\gamma \\ -c_\beta s_\alpha & c_\alpha s_\gamma + c_\gamma s_\alpha s_\beta & c_\alpha c_\gamma - s_\alpha s_\beta s_\gamma \end{bmatrix}$
$Z_\alpha Y_\beta Z_\gamma = \begin{bmatrix} c_\alpha c_\beta c_\gamma - s_\alpha s_\gamma & -c_\gamma s_\alpha - c_\alpha c_\beta s_\gamma & c_\alpha s_\beta \\ c_\alpha s_\gamma + c_\beta c_\gamma s_\alpha & c_\alpha c_\gamma - c_\beta s_\alpha s_\gamma & s_\alpha s_\beta \\ -c_\gamma s_\beta & s_\beta s_\gamma & c_\beta \end{bmatrix}$	$Z_\alpha Y_\beta X_\gamma = \begin{bmatrix} c_\alpha c_\beta & c_\alpha s_\beta s_\gamma - c_\gamma s_\alpha & s_\alpha s_\gamma + c_\alpha c_\gamma s_\beta \\ c_\beta s_\alpha & c_\alpha c_\gamma + s_\alpha s_\beta s_\gamma & c_\gamma s_\alpha s_\beta - c_\alpha s_\gamma \\ -s_\beta & c_\beta s_\gamma & c_\beta c_\gamma \end{bmatrix}$
$Z_\alpha X_\beta Z_\gamma = \begin{bmatrix} c_\alpha c_\gamma - c_\beta s_\alpha s_\gamma & -c_\alpha s_\gamma - c_\beta c_\gamma s_\alpha & s_\alpha s_\beta \\ c_\gamma s_\alpha + c_\alpha c_\beta s_\gamma & c_\alpha c_\beta c_\gamma - s_\alpha s_\gamma & -c_\alpha s_\beta \\ s_\beta s_\gamma & c_\gamma s_\beta & c_\beta \end{bmatrix}$	$Z_\alpha X_\beta Y_\gamma = \begin{bmatrix} c_\alpha c_\gamma - s_\alpha s_\beta s_\gamma & -c_\beta s_\alpha & c_\alpha s_\gamma + c_\gamma s_\alpha s_\beta \\ c_\gamma s_\alpha + c_\alpha s_\beta s_\gamma & c_\alpha c_\beta & s_\alpha s_\gamma - c_\alpha c_\gamma s_\beta \\ -c_\beta s_\gamma & s_\beta & c_\beta c_\gamma \end{bmatrix}$



Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

Driving datasets in a nutshell

All of your sensor modalities have vastly different rates and synchronizations



Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

123D: unifying open datasets

	Dataset	Year	Scale: Dur. [h]	Dist. [km]	Logs [#]
Manual	nuScenes [6]	2020	5.6	100.9	1,000
	WOD-Perc. [56]	2020	6.4	154.0	1,150
	AV2-Sens. [66]	2021	4.4	87.5	1,000
	PandaSet [67]	2021	0.2	8.3	103
	KITTI-360 [46]	2022	2.7	73.7	9
Auto-labeled	WOD-Mot. [21]	2021	574.1	10,323.5*	103,354
	nuPlan [35]	2024	1,174.3	17,808.6	15,910
	– mini		7.2	103.0	64
	PAI-AV [51]	2025	1,707.0	69,265.7	307,332
	– NCore [53]	2026	6.3	167.6	1,147
Synth.	CARLA [20]	2017	*var.*	*var.*	*var.*
	– L3AD [50]	2026	7.3	138.7	789

> 3,300 hours

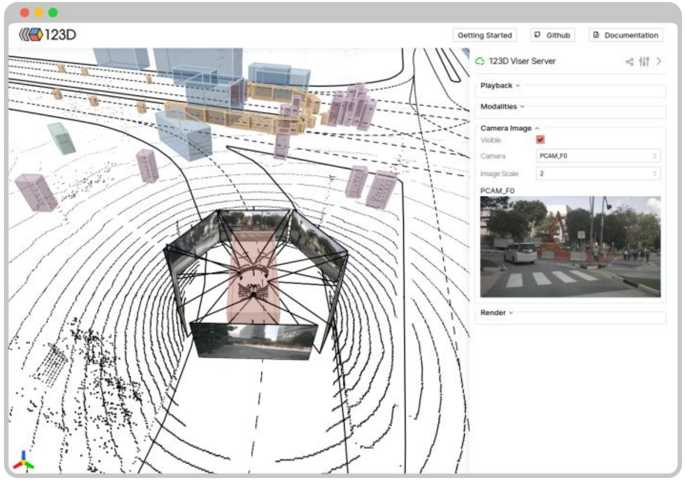
Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

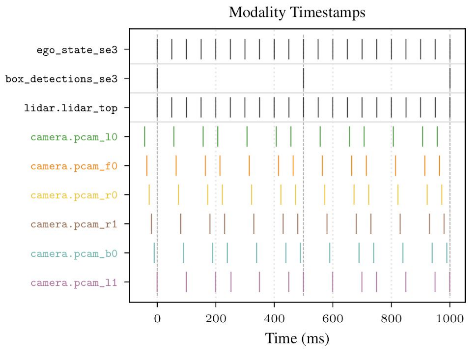
123D: unifying open datasets

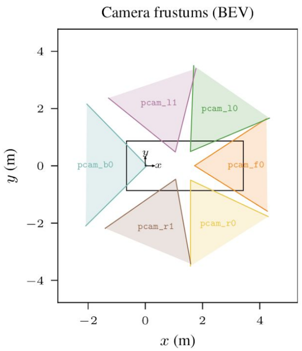
	Dataset	Year	Scale: Dur. [h]	Dist. [km]	Logs [#]
Manual	nuScenes [6]	2020	5.6	100.9	1,000
	WOD-Perc. [56]	2020	6.4	154.0	1,150
	AV2-Sens. [66]	2021	4.4	87.5	1,000
	PandaSet [67]	2021	0.2	8.3	103
	KITTI-360 [46]	2022	2.7	73.7	9
Auto-labeled	WOD-Mot. [21]	2021	574.1	10,323.5*	103,354
	nuPlan [35]	2024	1,174.3	17,808.6	15,910
	– mini		7.2	103.0	64
	PAI-AV [51]	2025	1,707.0	69,265.7	307,332
	– NCore [53]	2026	6.3	167.6	1,147
Synth.	CARLA [20]	2017	*var.*	*var.*	*var.*
	– L3AD [50]	2026	7.3	138.7	789

> 3,300 hours

nuScenes:







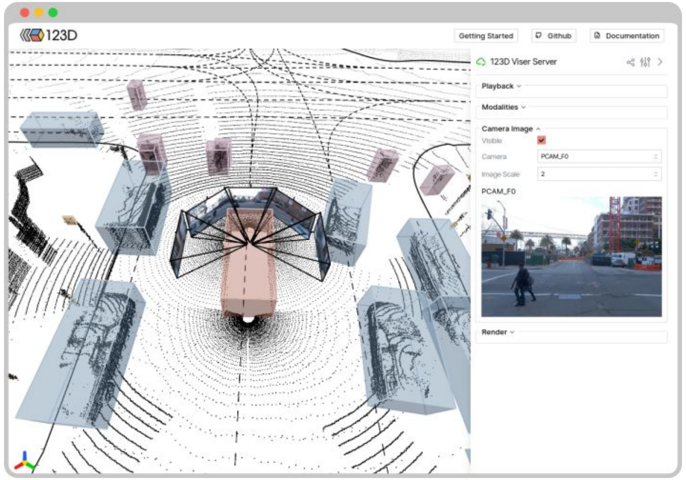
Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

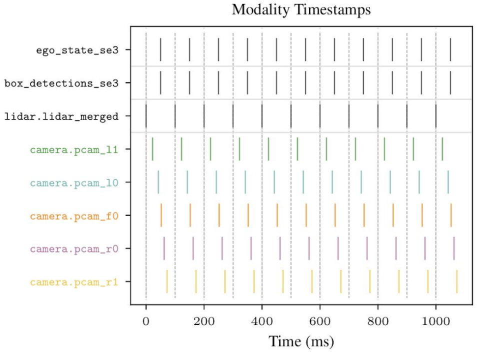
123D: unifying open datasets

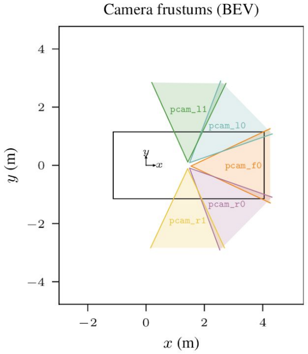
	Dataset	Year	Scale: Dur. [h]	Scale: Dist. [km]	Scale: Logs [#]
Manual	nuScenes [6]	2020	5.6	100.9	1,000
	WOD-Perc. [56]	2020	6.4	154.0	1,150
	AV2-Sens. [66]	2021	4.4	87.5	1,000
	PandaSet [67]	2021	0.2	8.3	103
	KITTI-360 [46]	2022	2.7	73.7	9
Auto-labeled	WOD-Mot. [21]	2021	574.1	10,323.5*	103,354
	nuPlan [35]	2024	1,174.3	17,808.6	15,910
	– mini		7.2	103.0	64
	PAI-AV [51]	2025	1,707.0	69,265.7	307,332
	– NCore [53]	2026	6.3	167.6	1,147
Synth.	CARLA [20]	2017	var.	var.	var.
	– L3AD [50]	2026	7.3	138.7	789

$\underbrace{\hspace{10em}}_{> 3{,}300 \text{ hours}}$

Waymo Open Dataset







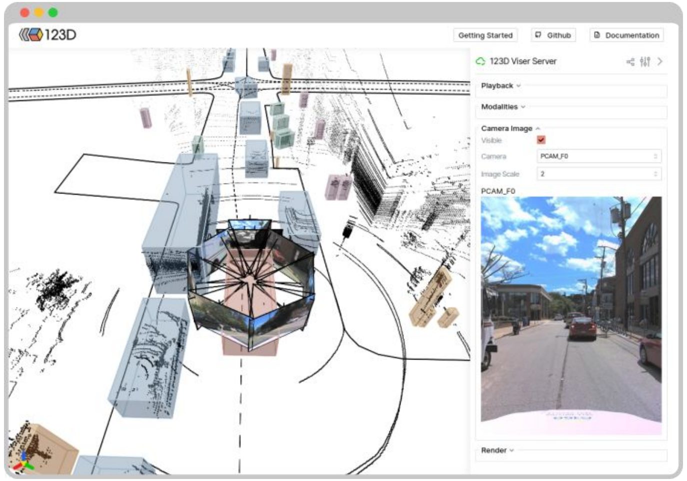
Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

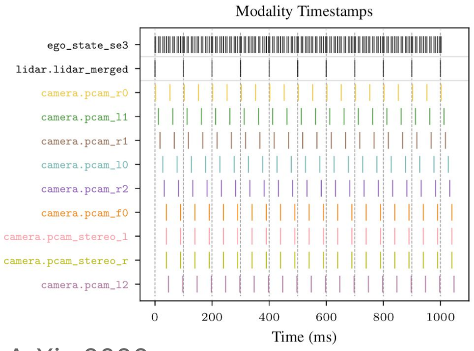
123D: unifying open datasets

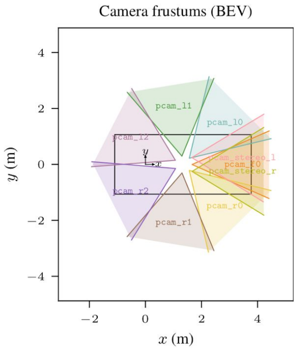
	Dataset	Year	Scale: Dur. [h]	Scale: Dist. [km]	Scale: Logs [#]
Manual	nuScenes [6]	2020	5.6	100.9	1,000
Manual	WOD-Perc. [56]	2020	6.4	154.0	1,150
Manual	**AV2-Sens. [66]**	**2021**	**4.4**	**87.5**	**1,000**
Manual	PandaSet [67]	2021	0.2	8.3	103
Manual	KITTI-360 [46]	2022	2.7	73.7	9
Auto-labeled	WOD-Mot. [21]	2021	574.1	10,323.5*	103,354
Auto-labeled	nuPlan [35]	2024	1,174.3	17,808.6	15,910
Auto-labeled	– mini	2024	7.2	103.0	64
Auto-labeled	PAI-AV [51]	2025	1,707.0	69,265.7	307,332
Auto-labeled	– NCore [53]	2026	6.3	167.6	1,147
Synth.	CARLA [20]	2017	*var.*	*var.*	*var.*
Synth.	– L3AD [50]	2026	7.3	138.7	789

$\underbrace{\hphantom{xxxxxxxxxxxxxxxx}}_{> 3{,}300 \text{ hours}}$

Argoverse 2:







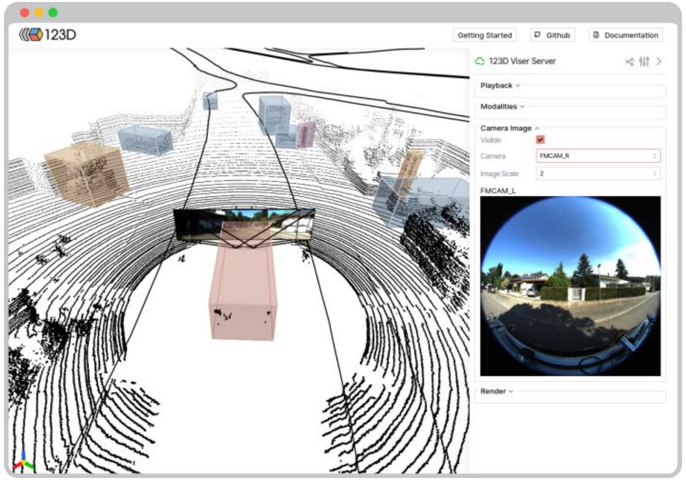
Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

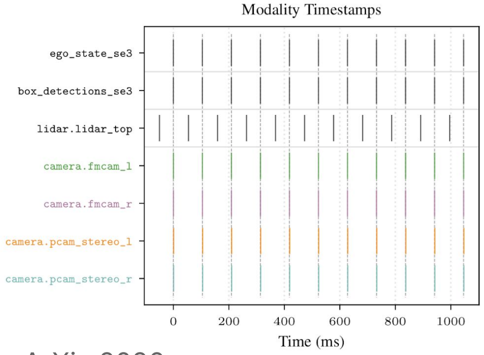
123D: unifying open datasets

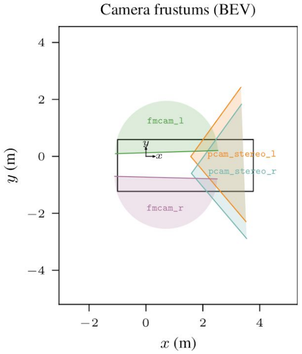
	Dataset	Year	Scale: Dur. [h]	Dist. [km]	Logs [#]
Manual	nuScenes [6]	2020	5.6	100.9	1,000
	WOD-Perc. [56]	2020	6.4	154.0	1,150
	AV2-Sens. [66]	2021	4.4	87.5	1,000
	PandaSet [67]	2021	0.2	8.3	103
	KITTI-360 [46]	**2022**	**2.7**	**73.7**	**9**
Auto-labeled	WOD-Mot. [21]	2021	574.1	10,323.5*	103,354
	nuPlan [35]	2024	1,174.3	17,808.6	15,910
	– mini		7.2	103.0	64
	PAI-AV [51]	2025	1,707.0	69,265.7	307,332
	– NCore [53]	2026	6.3	167.6	1,147
Synth.	CARLA [20]	2017	*var.*	*var.*	*var.*
	– L3AD [50]	2026	7.3	138.7	789

> 3,300 hours

KITTI-360:







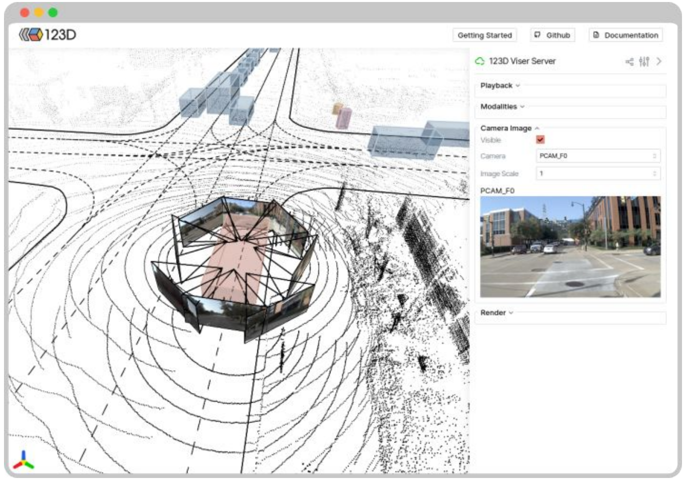
Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

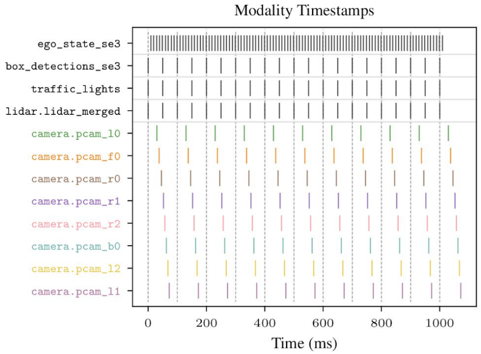
123D: unifying open datasets

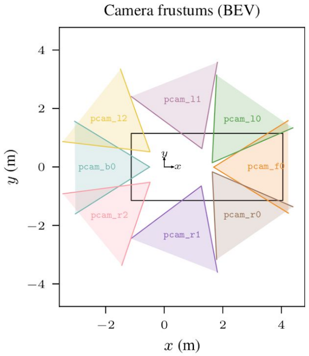
	Dataset	Year	Scale: Dur. [h]	Scale: Dist. [km]	Scale: Logs [#]
Manual	nuScenes [6]	2020	5.6	100.9	1,000
	WOD-Perc. [56]	2020	6.4	154.0	1,150
	AV2-Sens. [66]	2021	4.4	87.5	1,000
	PandaSet [67]	2021	0.2	8.3	103
	KITTI-360 [46]	2022	2.7	73.7	9
Auto-labeled	WOD-Mot. [21]	2021	574.1	10,323.5*	103,354
	nuPlan [35]	**2024**	**1,174.3**	**17,808.6**	**15,910**
	– mini		**7.2**	**103.0**	**64**
	PAI-AV [51]	2025	1,707.0	69,265.7	307,332
	– NCore [53]	2026	6.3	167.6	1,147
Synth.	CARLA [20]	2017	*var.*	*var.*	*var.*
	– L3AD [50]	2026	7.3	138.7	789

$\underbrace{\hspace{10em}}_{> 3{,}300 \text{ hours}}$

nuPlan:







Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

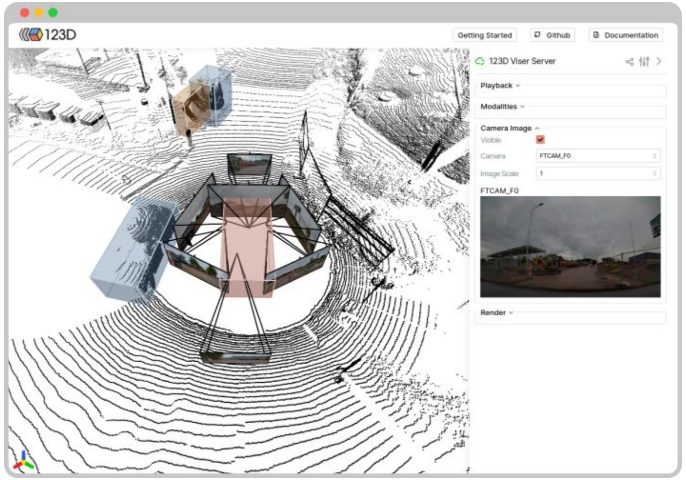
123D: unifying open datasets

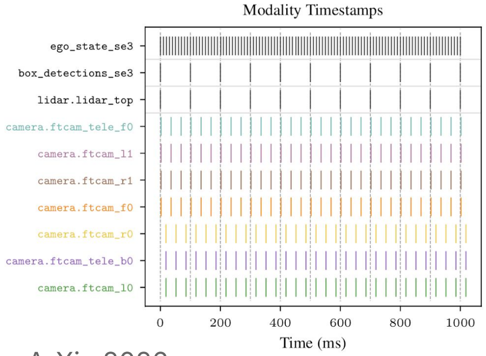
	Dataset	Year	Dur. [h]	Dist. [km]	Logs [#]
Manual	nuScenes [6]	2020	5.6	100.9	1,000
	WOD-Perc. [56]	2020	6.4	154.0	1,150
	AV2-Sens. [66]	2021	4.4	87.5	1,000
	PandaSet [67]	2021	0.2	8.3	103
	KITTI-360 [46]	2022	2.7	73.7	9
Auto-labeled	WOD-Mot. [21]	2021	574.1	10,323.5*	103,354
	nuPlan [35]	2024	1,174.3	17,808.6	15,910
	– mini	2024	7.2	103.0	64
	PAI-AV [51]	**2025**	**1,707.0**	**69,265.7**	**307,332**
	– NCore [53]	**2026**	**6.3**	**167.6**	**1,147**
Synth.	CARLA [20]	2017	*var.*	*var.*	*var.*
	– L3AD [50]	2026	7.3	138.7	789

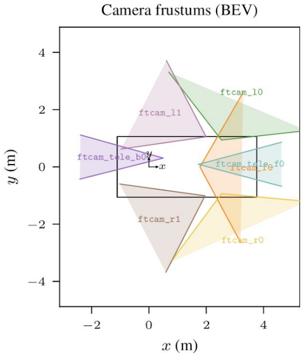
Dur. [h], Dist. [km] and Logs [#] are grouped under "Scale".

$\underbrace{\hspace{10em}}_{}$ > 3,300 hours

NVIDIA PhysicalAI-AV:





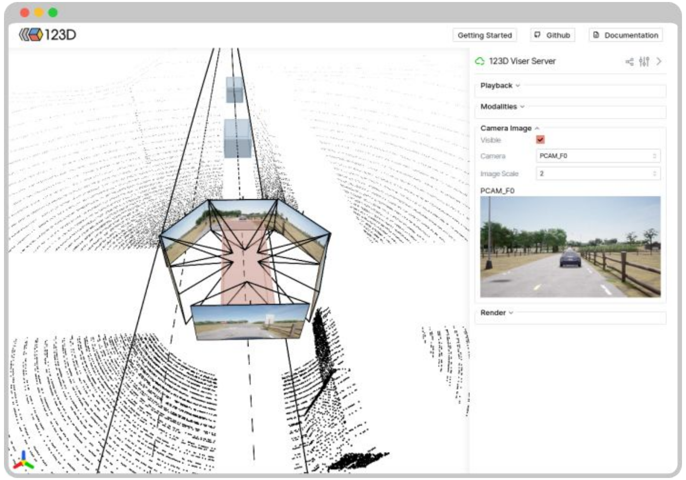


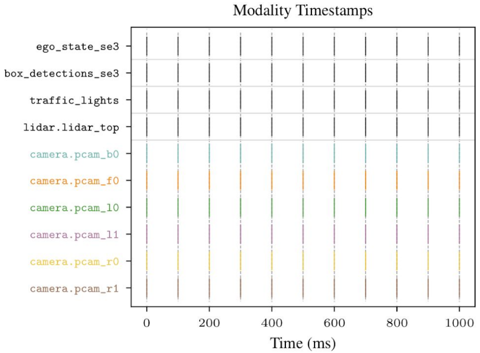
123D: unifying open datasets

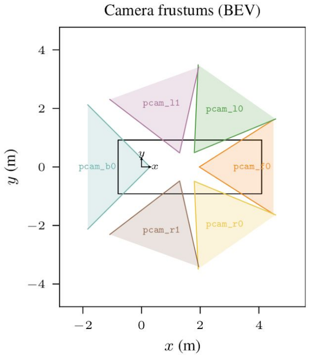
	Dataset	Year	Scale: Dur. [h]	Scale: Dist. [km]	Scale: Logs [#]
Manual	nuScenes [6]	2020	5.6	100.9	1,000
Manual	WOD-Perc. [56]	2020	6.4	154.0	1,150
Manual	AV2-Sens. [66]	2021	4.4	87.5	1,000
Manual	PandaSet [67]	2021	0.2	8.3	103
Manual	KITTI-360 [46]	2022	2.7	73.7	9
Auto-labeled	WOD-Mot. [21]	2021	574.1	10,323.5*	103,354
Auto-labeled	nuPlan [35]	2024	1,174.3	17,808.6	15,910
Auto-labeled	– mini	2024	7.2	103.0	64
Auto-labeled	PAI-AV [51]	2025	1,707.0	69,265.7	307,332
Auto-labeled	– NCore [53]	2026	6.3	167.6	1,147
Synth.	CARLA [20]	2017	*var.*	*var.*	*var.*
Synth.	– L3AD [50]	2026	7.3	138.7	789

> 3,300 hours

CARLA / L3AD:







Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

Pipeline

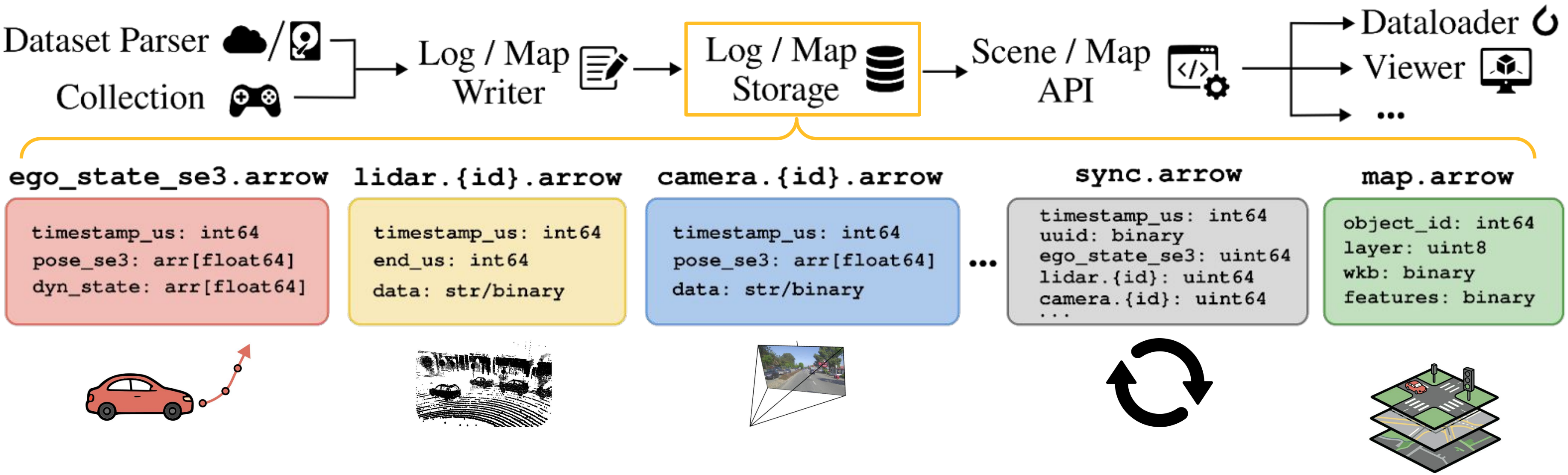
Parse existing dataset
from cloud/ local storage

```
py123d-conversion dataset=...
```

... or directly collect in simulation

Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

Pipeline



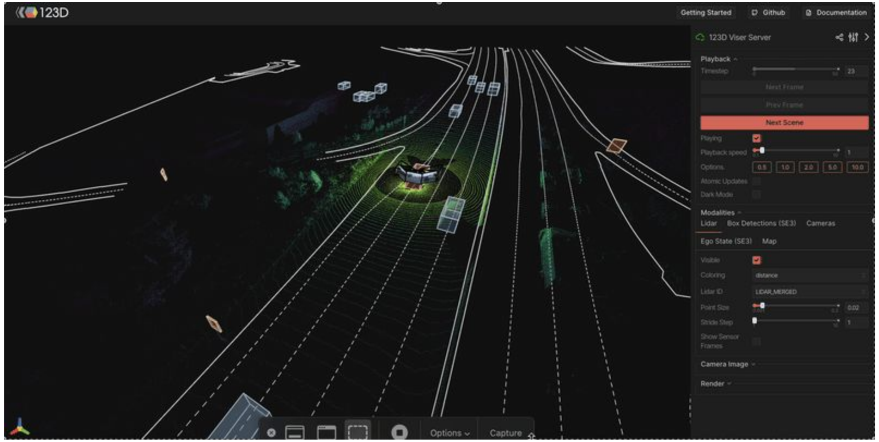
Independent timestamped stream per modality
+ Pre-computed synchronization table

Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

Pipeline


```python
from py123d.api import MapAPI, SceneAPI, SceneFilter, get_filtered_scenes
scene_filter = SceneFilter( # Mix datasets with different rigs and native rates.
    split_names=["av2-sensor_train", "nuscenes_train", "wod-perception_train"],
    target_iteration_duration_s=0.5, # 2 Hz shared reference rate
    future_duration_s=4.0, history_duration_s=1.0, shuffle=True
)
scene: SceneAPI = get_filtered_scenes(scene_filter, data_root=...)[0]
# 1. Sync access: ego / lidar at initial iteration.
ego = scene.get_ego_state_se3_at_iteration(iteration=0)
lidar = scene.get_lidar_at_iteration(iteration=0, lidar_id="lidar_top")
# 2. Async access: the nearest front-camera frame to the lidar sweep.
camera = scene.get_camera_at_timestamp(
    timestamp=lidar.timestamp_start, camera_id="pcam_f0", criteria="nearest",
)
# 3. Map access: query nearby lanes and crosswalks around ego.
map_api: MapAPI = scene.get_map_api()
nearby = map_api.get_map_objects_in_radius(
    point=ego.center_3d, radius=50.0, layers=["lane", "crosswalk"],
)
```

Python API



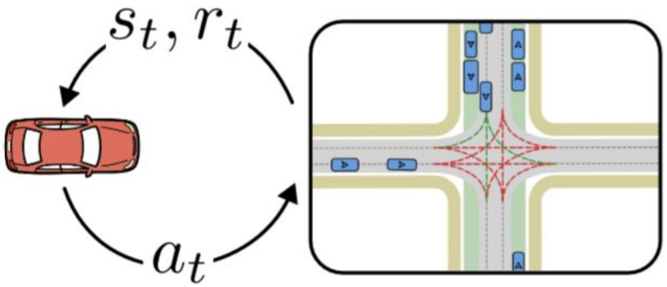
3D Viewer

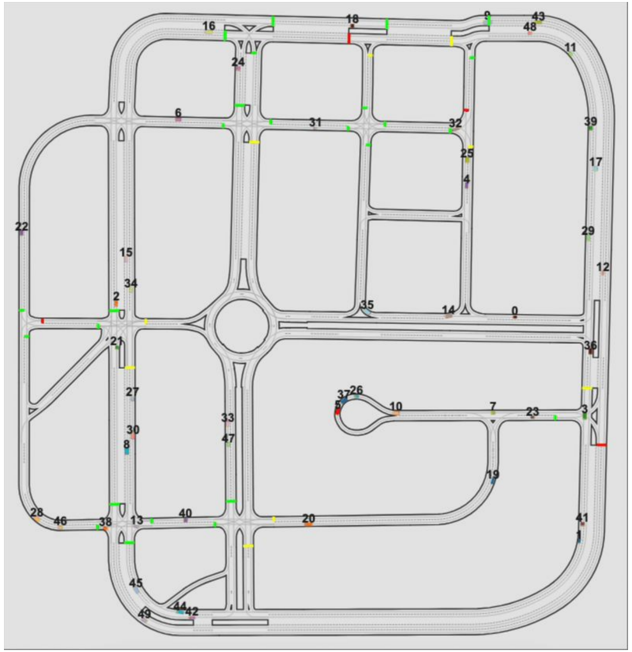
Dauner et al., 123D: Unifying Multi-Modal Autonomous Driving Data at Scale. ArXiv 2026

Open Simulators

123Drive + PufferDrive

High Throughput Reinforcement Learning

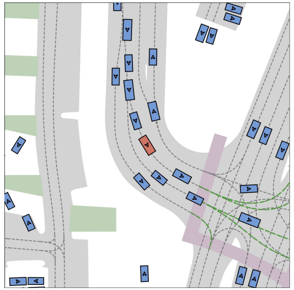




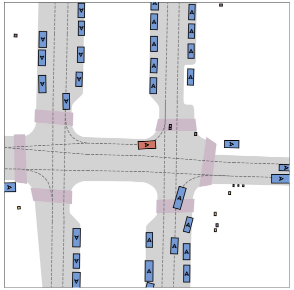
github.com/Emerge-Lab/PufferDrive
github.com/vcharraut/123Drive

Cornelisse et al., PufferDrive: A Fast and Friendly Driving Simulator for Training and Evaluating RL Agents.

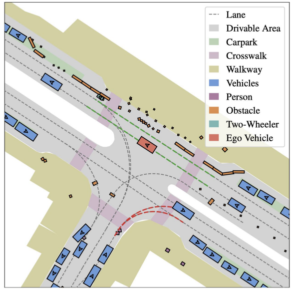
123Drive + PufferDrive



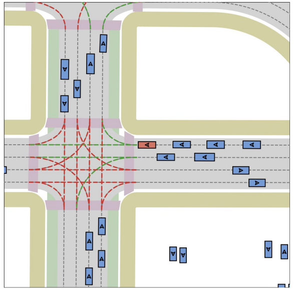
WOD-Motion



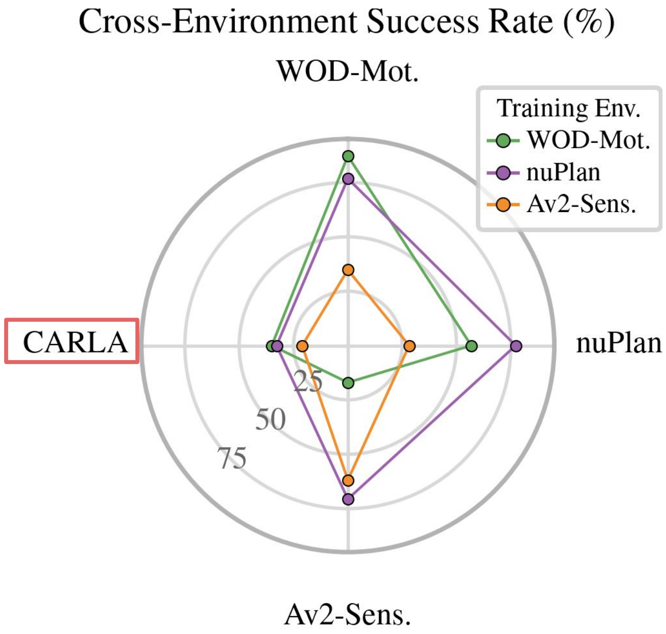
Av2-Sensor



nuPlan



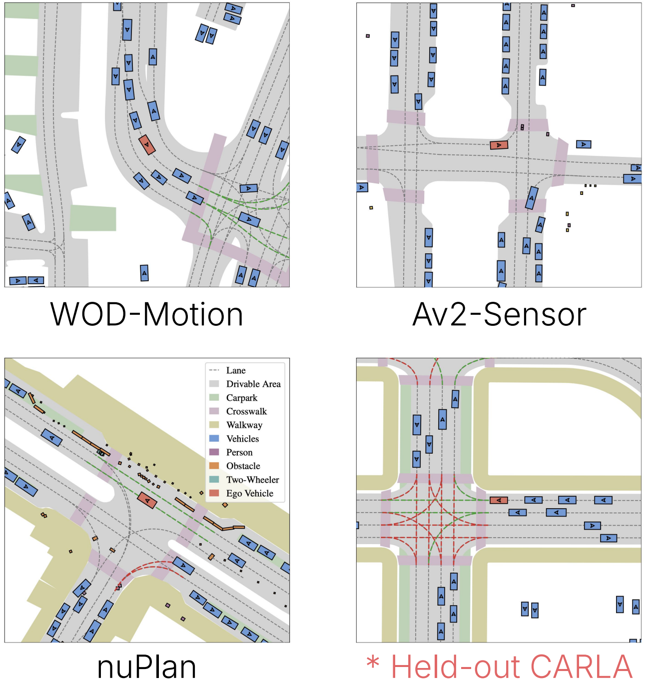
* Held-out CARLA

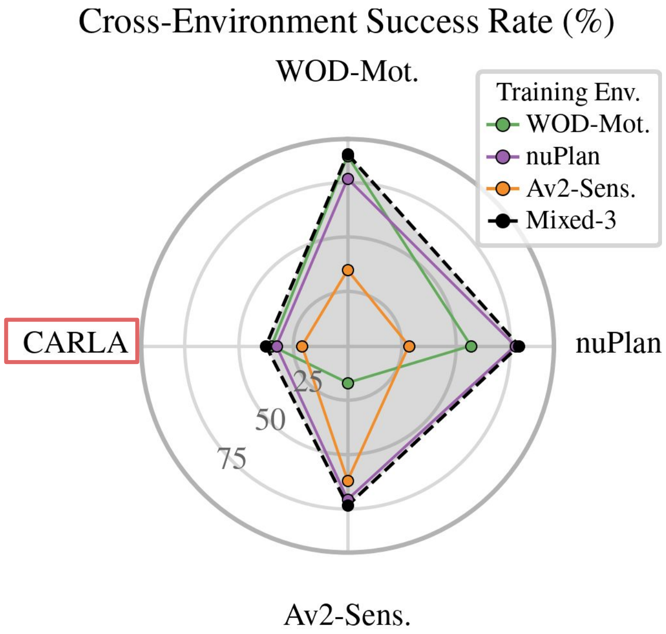


In-domain evaluated policies perform best in-distribution

Cornelisse et al., PufferDrive: A Fast and Friendly Driving Simulator for Training and Evaluating RL Agents.

123Drive + PufferDrive





Simple mixture generalizes across observed environments

Cornelisse et al., PufferDrive: A Fast and Friendly Driving Simulator for Training and Evaluating RL Agents.

AlpaSim Challenge 2026

- **AlpaSim**: open-source, modular, photorealistic closed-loop simulation
- Large-scale **held-out** evaluation environments for leaderboard

PAI-AV Towards Frontier Scale

New dataset (1,700 hours)
Raw real-world driving data without heavy curation

Geographically diverse
25 countries, 2500+ cities

Unconstrained private data & model use
Subject to licensing, disclosure & integrity rules

nuPlan Enriching Existing Ecosystem

Established dataset (120 hours)
Open annotations, maps, and curated splits

Lower entry barrier
Hundreds of community baselines

Only open-source models & tools permitted
Fully reproducible competition entries

Cao et al., AlpaSim: A Modular, Lightweight, and Data-Driven Research Simulator for Autonomous Driving.

PAI-AV + AlpaSim

Cao et al., AlpaSim: A Modular, Lightweight, and Data-Driven Research Simulator for Autonomous Driving.

nuPlan + AlpaSim

Cao et al., AlpaSim: A Modular, Lightweight, and Data-Driven Research Simulator for Autonomous Driving.

KE:SAI

Open Science Autonomy Lab

Thank You!

kesai.eu/join