The Data Cards Playbook

Datacard 6

Synthetic data set generated using Simcenter Prescan - Vulnerable Road Users in urban driving scenario.
Data_Set_6 (for consortium members)
Synthetic data set generated using Simcenter Prescan - Vulnerable Road Users in urban driving scenario (publicly available data set)

Synthetic data set generated by Siemens Industry Software Netherlands B.V., which contains a simple scenario: pedestrian crossing in front of a vehicle. The data set in an extended KITTI data format, since it contains unique object IDs and Simcenter Prescan object categories, in addition to the standard KITTI data format. The data set contains image files as well as point cloud data files, radar data, as well ideal depth map, recorded at each timestamp.

DATASET LINK

Dataset Link

DATA PUBLISHER(S)

Siemens Industry Software Netherlands B.V., RTD Department,

Alexandru Forrai, alexandru.forrai@siemens.com

Funding Sources (Optional)

INSTITUTION(S)	FUNDING OR GRANT SUMMARY(IES)
Funded by the European Union	The AIthena project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement number 101076754. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or CINEA. Neither the European Union nor the granting authority can be held responsible for them.

Dataset Overview

DATA SUBJECT(S)	CONTENT DESCRIPTION
[ ] Sensitive Data about people [x] Non-Sensitive Data about people [ ] Data about natural phenomena [ ] Data about places and objects [x] Synthetically generated data [ ] Data about systems or products and their behaviours [ ] Unknown [ ] Others (Please specify):	**Additional Notes:** synthetic data set, considering a simple scenario, pedestrian crossing in front of a vehicle, normal weather, and illumination conditions. The data set might be suitable to develop (test) algorithms and data pipelines related to multiple target tracking, considering occlusion and truncation as well as sensor fusion based on camera and lidar data.

DATASET SNAPSHOT

Size of Dataset	175MB
Number of Instances	849 files, 3 folders
Number of Fields	123456
Labelled Classes	Car, Van, Truck, Pedestrian, Person_Sitting, Cyclist, Tram, DontCare
Number of Labels	9
Average Labels Per Instance	10
Algorithmic Labels	Ground truth
Human Labels	2
Other Characteristics	None

Scene Descriptors and ODD

Labels and percentage (if applicable)

- *Weather*
  - *'Clear', 'Rain', 'Light Fog', 'Dense Fog', 'Snow and Light Fog'*
- *Daylight*
  - *'Day', 'Night'*
- *Collision - None*
- *Overtaking - None*
- *Localisation – Ground truth*
- *Environment*
  - *Suburban, urban*

Sensitivity of Data

SENSITIVITY TYPE(S)	FIELD(S) WITH SENSITIVE DATA	SECURITY AND PRIVACY HANDLING
None	**Intentionally Collected Sensitive Data -** None **Unintentionally Collected Sensitive Data -** None **Additional Notes:** None	**Additional Notes:** the data set is generated for AITHENA project consortium members. A small data set is also available at: Synthetic data set generated using Simcenter Prescan - Vulnerable Road Users in urban driving scenario

Dataset Version and Maintenance

MAINTENANCE STATUS	VERSION DETAILS	MAINTENANCE PLAN

Regularly Updated (New versions of the dataset have been or will continue to be made available.)	**Current Version:** 1.0 **Last Updated:** Aug/2025 **Release Date:** Aug/2025	**Versioning:** Versioning via team, when a new data set is released a new version number will be associated **Updates:** the data has been extended with radar data and ideal depth map. **Errors:** None **Feedback:** None **Additional Notes:** None

Example of Data Points

PRIMARY DATA MODALITY	EXAMPLE OF DATAPOINT
Image Data (yes) Text Data (yes) Tabular Data (yes) Audio Data (no) Video Data (yes) Time Series (yes) Graph Data (no) Geospatial Data (no) Multimodal (Please Specify) (none) Unknown Others (Please specify)	

Values	Name	Description
1	object ID	Unique object ID from Prescan
1	object type	Describes the type of the object based on Simcenter Prescan object types

```
#Values    Name      Description
---------------------------------------------------------------------------
   1    type         Describes the type of object: 'Car', 'Van', 'Truck',
                     'Pedestrian', 'Person_sitting', 'Cyclist', 'Tram',
                     'Misc' or 'DontCare'
   1    truncated    Float from 0 (non-truncated) to 1 (truncated), where
                     truncated refers to the object leaving image boundaries
   1    occluded     Integer (0,1,2,3) indicating occlusion state:
                     0 = fully visible, 1 = partly occluded
                     2 = largely occluded, 3 = unknown
   1    alpha        Observation angle of object, ranging [-pi..pi]
   4    bbox         2D bounding box of object in the image (0-based index):
                     contains left, top, right, bottom pixel coordinates
   3    dimensions   3D object dimensions: height, width, length (in meters)
   3    location     3D object location x,y,z in camera coordinates (in meters)
   1    rotation_y   Rotation ry around Y-axis in camera coordinates [-pi..pi]
   1    score        Only for results: Float, indicating confidence in
                     detection, needed for p/r curves, higher is better.
```

Access, Retention, & Wipeout

Access

ACCESS TYPE	DOCUMENTATION LINK(S)
External - Open	*Synthetic data set generated using Simcenter Prescan - Vulnerable Road Users in urban driving scenario.* *SYNTHETIC DATA DESCRIPTION - Data generated using Simcenter Prescan*

Provenance – synthetically generated data

Collection

METHODOLOGY DETAIL(S)	SOURCE DESCRIPTION(S)	DATA PROCESSING
Collection Type: **Source:** Synthetically generated data, using Simcenter Prescan. **Is this source considered sensitive or high-risk?** No **Dates of Collection:** [Aug 2023 - Aug 2025] **Primary modality of collected data:** Time Series **Update Frequency for collected data:** Quarterly	**Source:** Scenario defined in Simcenter Prescan **Additional Notes:** None	**Collection Method or Source** **Description:** Automated data recording using Simcenter Prescan and Matlab (MathWorks). **Methods employed:** Data is saved in extended KITTI data format. **Tools or libraries:** Simcenter Prescan and Matlab (MathWorks). **Additional Notes:** None.
COLLECTION PROCESS		
Synthetic data recording according to the sensors data collection frequency.		

Human and Other Sensitive Attributes

SENSITIVE HUMAN ATTRIBUTE(S)	REPRESENTED GROUPS OF PEOPLE
[x] Gender	[ ] Non-disabled adult
[ ] Socio-economic status	[ ] Non-disabled child
[ ] Geography	[ ] Non-disabled elderly
[ ] Language	[ ] Disabled adult
[x] Age	[ ] Disabled child
[x] Culture	[ ] Disabled elderly
[ ] Others (Please Specify):	

Annotations & Labelling

ANNOTATION WORKFORCE TYPE	ANNOTATION DISTRIBUTION(S)	ANNOTATION MEANS DESCRIPTION(S)
Machine-Generated Annotations (x)	Annotations type machine generated [100%]	**Description:** data is annotated using the ground truth information from the simulation engine – Simcenter Prescan. **Link:** none **Platforms, tools, or libraries:** Simcenter Prescan **Additional Notes:** None

Validation Methods

Did you:
[x] Check the labels are correct
[x] Check the capture (image, text) is correct
[x] Check the format is the expected
[ ] Other:

Known Applications & Benchmarks

ML APPLICATION(S)	ARCHITECTURE	EVALUATION
None	**(Model Name/type)** Not applicable **Model Card:** Not applicable **Additional Notes:** None	**(Model Name)** Note applicable

KNOWN CAVEATS

Synthetic data set generated using Simcenter Prescan - Vulnerable Road Users in urban driving scenario.
Expected Performance: Synthetic data set, high performance in terms of visualization

Known Caveats: None

Additional Notes: None

Reflections on Data

Data Usage

1. Develop and test multiple target tracking algorithms
2. Develop and test sensor fusion algorithm using camera, radar and lidar sensors.
3. Develop and test monocular depth estimation algorithms