KI ABSICHERUNG

Safe AI for Automated Driving

11th March 2021, Online, Interim Presentation

Generating synthetic learning, testing and validation data

Dr. Thomas Stauner, BMW AG

Why are we using synthetic data?

- ML relies on a vast amount of data. Data collection in the real world is not reproduceable. It can hardly support the study of specific effects.
- Synthetic data generation allows us to specify exactly what we want, to control influence factors and introduce targeted variations.

Same sensor position, different scene layouts

Why are we using synthetic data?

Why are we using synthetic data?

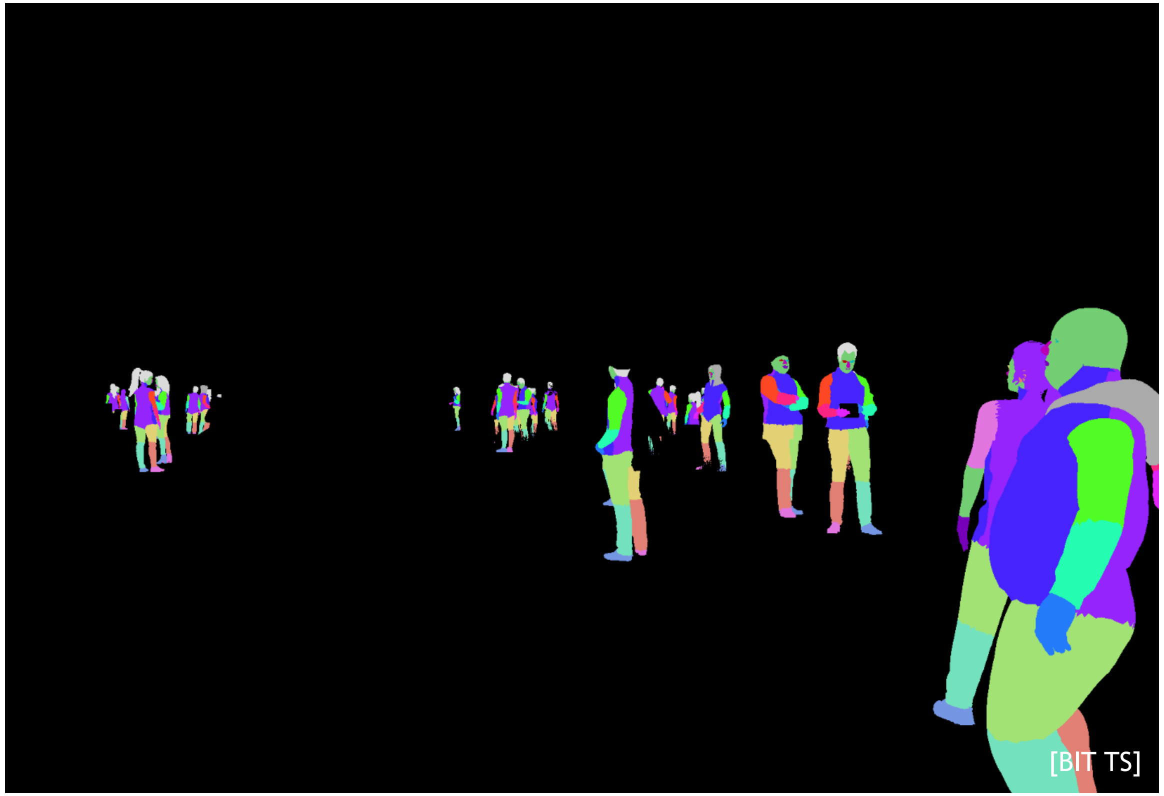
Extensive ground truth can be automatically computed from synthetic data

2d bounding boxes

3d bounding boxes

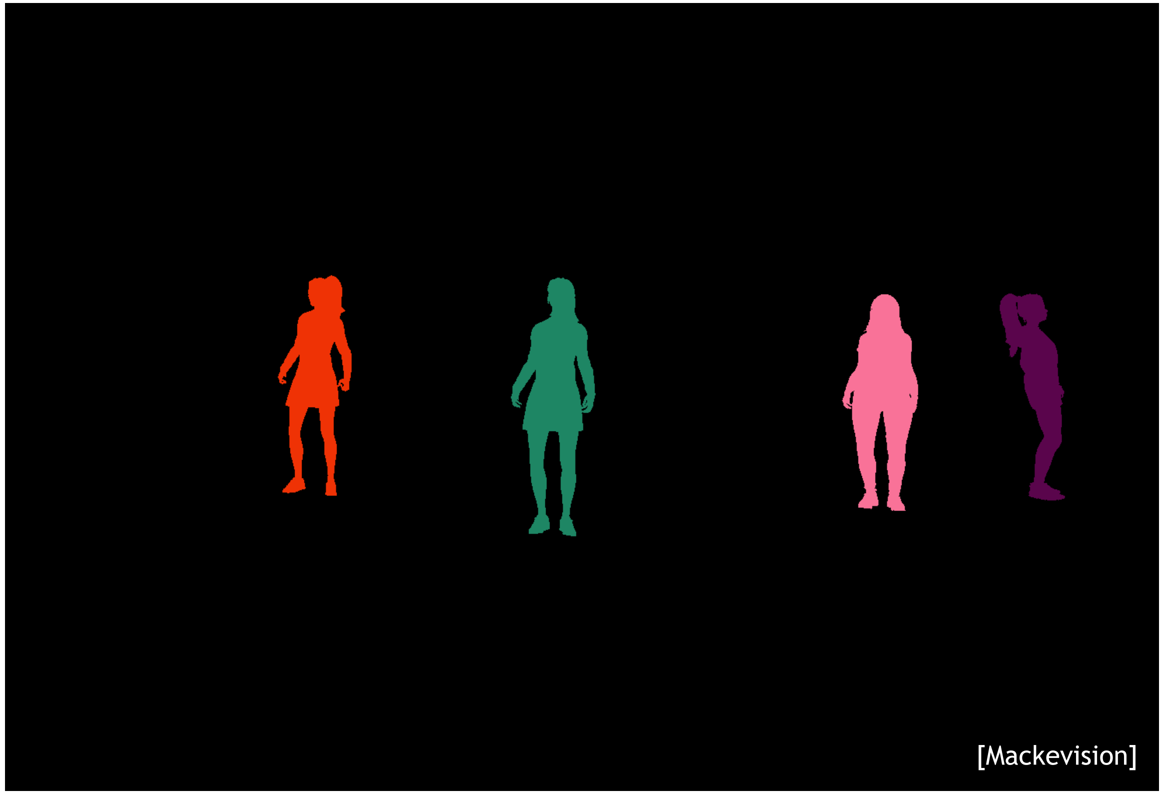
Extensive ground truth can be automatically computed from synthetic data

Semantic segmentation

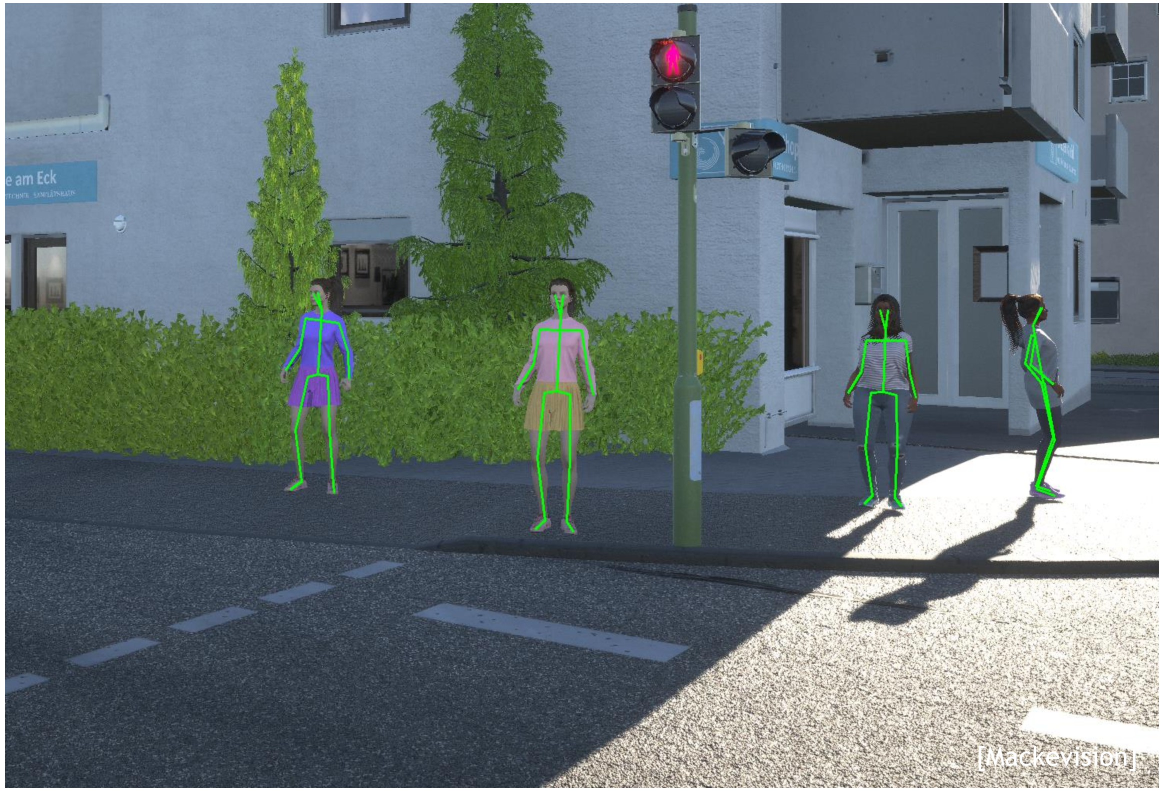


Bodypart segmentation

Extensive ground truth can be automatically computed from synthetic data



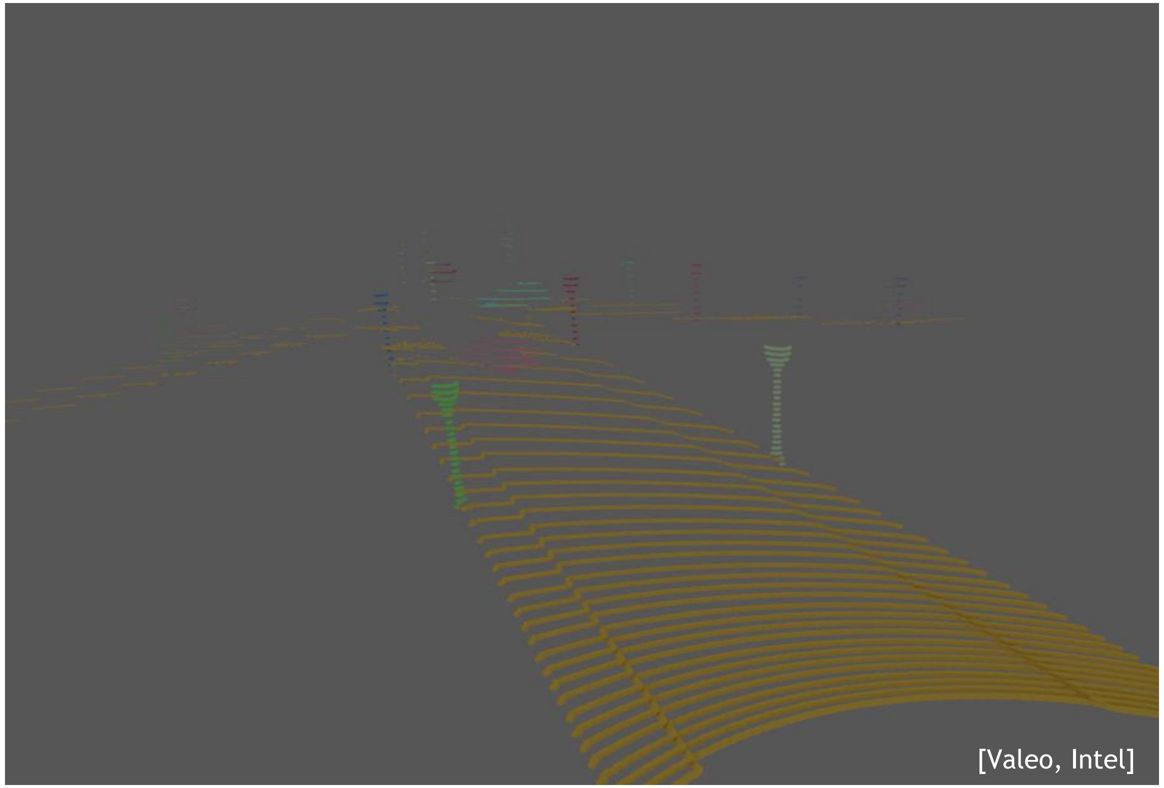
Instance segmentation



Pose data

Extensive ground truth can be automatically computed from synthetic data

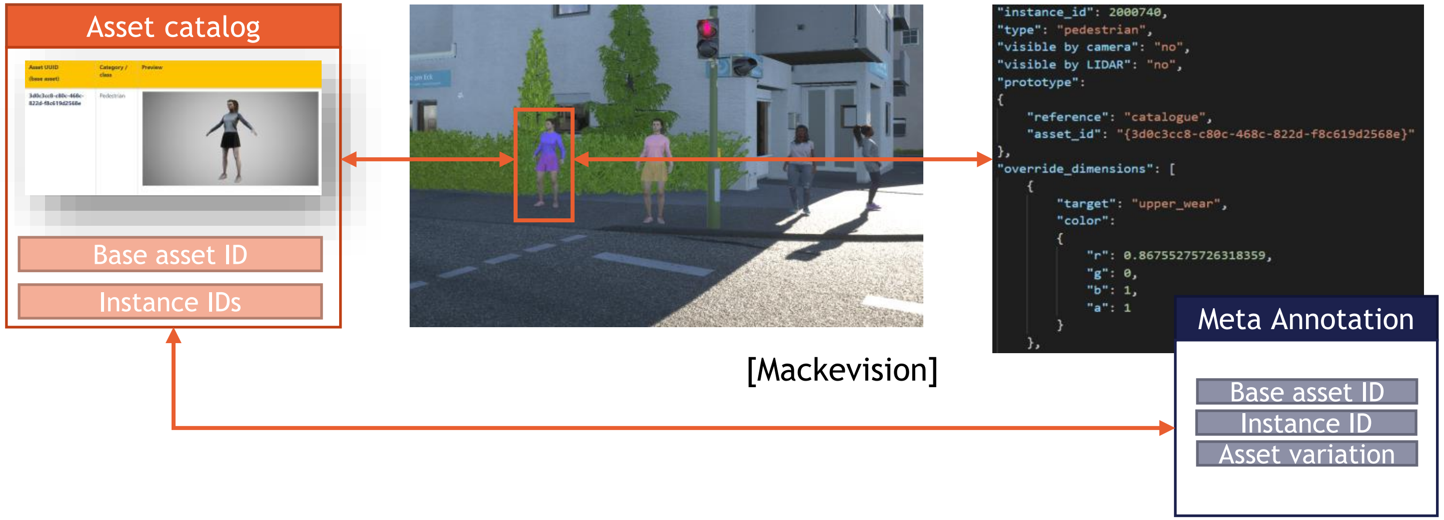
Depth map



Lidar instance segmentation (WIP)

For safety analysis, diverse meta data for synthetic images can be computed

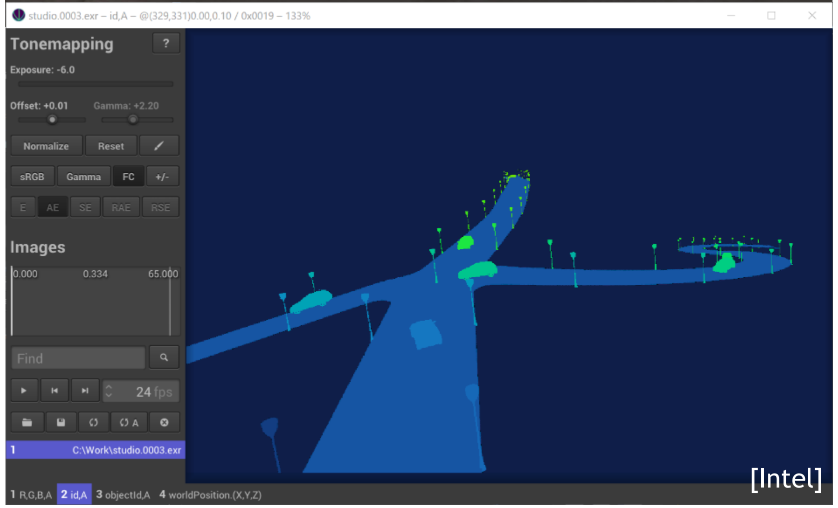
- For the systematic analysis of weaknesses of an AI function, rich meta information is required.
- It allows the engineer to retrieve semantic information w.r.t. an ontology for the situation depicted in a frame. Examples are body size of pedestrians or clothing colors.
- The implementation of such meta information requires considerable one-time effort.



[Mackevision]

Halftime in data production

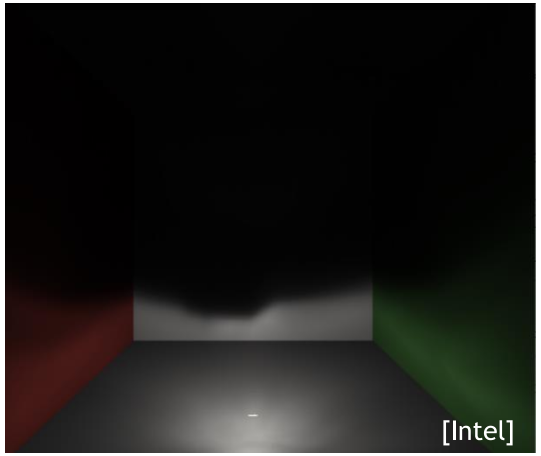
- ~100K frames were produced in 4 delivery tranches, with increasing complexity and increasing amount of annotations and metadata.
- A new data production toolchain with support for sensor models has been prepared:
  - Integration of Intel OSPRay Studio into BIT TS pipeline
    - Open glTF-based interface
    - Support for native materials
    - Support for animation, skinning, HDRI
    - Rich groundtruth + metadata
    - Valeo LIDAR plugin
      - PoC, 3 echos per point
      - WIP: tests, validation

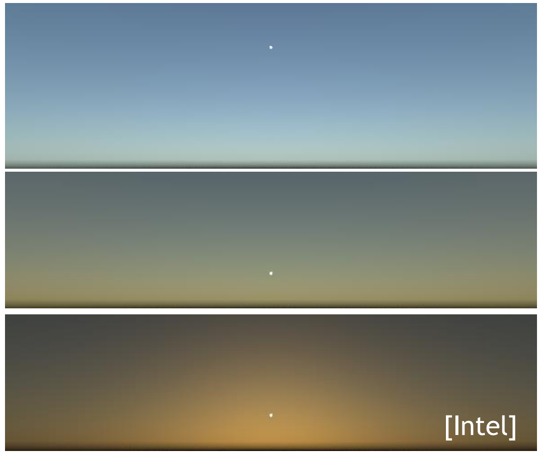


Completed extensions of OSPRay and error generator for future image features

- Measured light sources
- Sun-sky-illumination model
- Light: radiometric quantities
- Pixel filters
- Optimizations and bugfixes (e.g. HDRI poles)

- Error generator for
  - Contrast, distortion,
  - chromatic aberration, noise, blurring





Quality evaluation is an essential part of synthetic data generation

- **Goal: evaluate synthetic data w.r.t.**

Completeness

Does the data cover the relevant part of the world?

Correctness

Is the data plausible w.r.t. the ODD (operational design domain)?

Relevance

How do data variations affect training and inference performance?

Extended goals

Transferability of the quality evaluation to other data sets
Methods to bridge the domain gap

Prerequisites

- Availability of a notion of completeness
- Methods to measure coverage
- Methods to assess effect of coverage gaps

- Knowledge of the ground truth -> calibration
- Measurability of “generated” and “subjective” ground truth
- Methods to measure tool-induced deviations

- Measurement of training success w.r.t. reference data set
- Measurement of detection probability

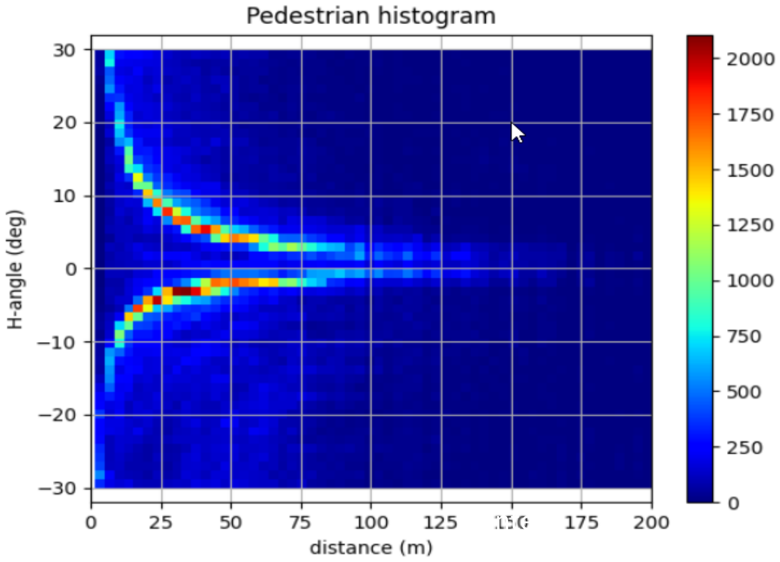
- Availability of homogenous meta information
- Automated post-processing tools
- Application of domain adaption methods to bridge domain gaps

Quality evaluation example: a data set needs to be evaluated as a whole

- Quality aspect: pedestrian position

30+pedestrians hidden in the single image.

- Measurement and evaluation



Distribution histogram of an example data set showing unexpected sharpness.

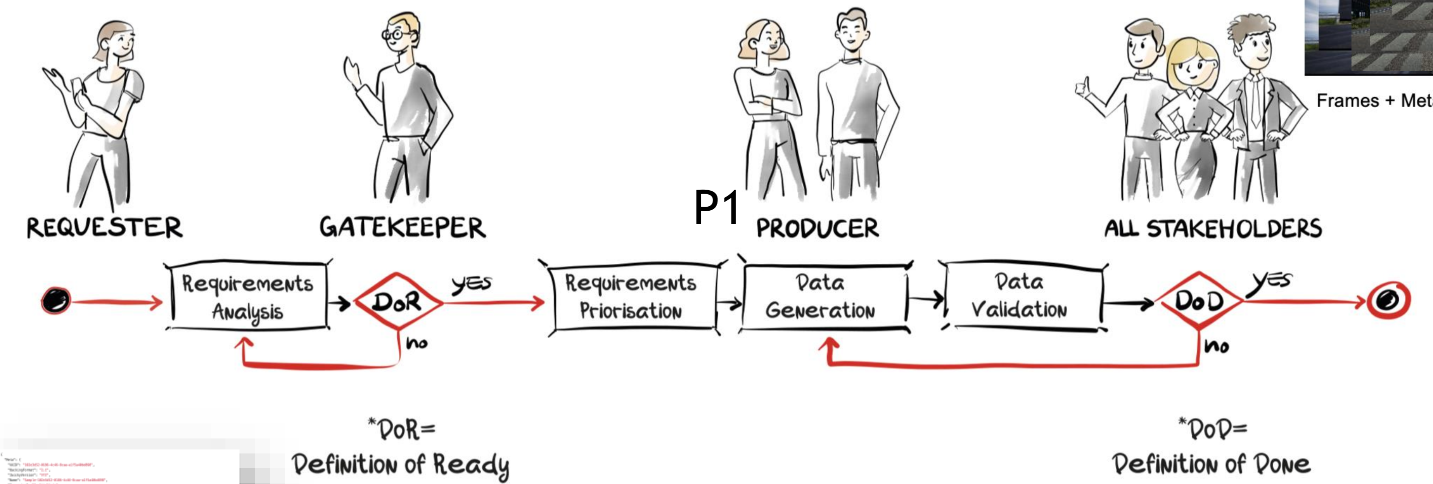
Requirements management process for diverse requirements established

Requirements for AI development

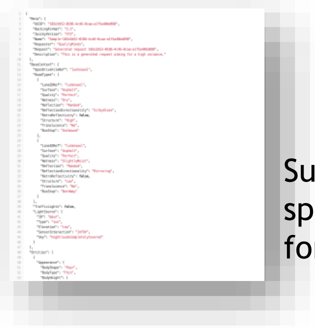
Requirements for safety analysis

Requirements on optical quality

Tech. requirements from data production

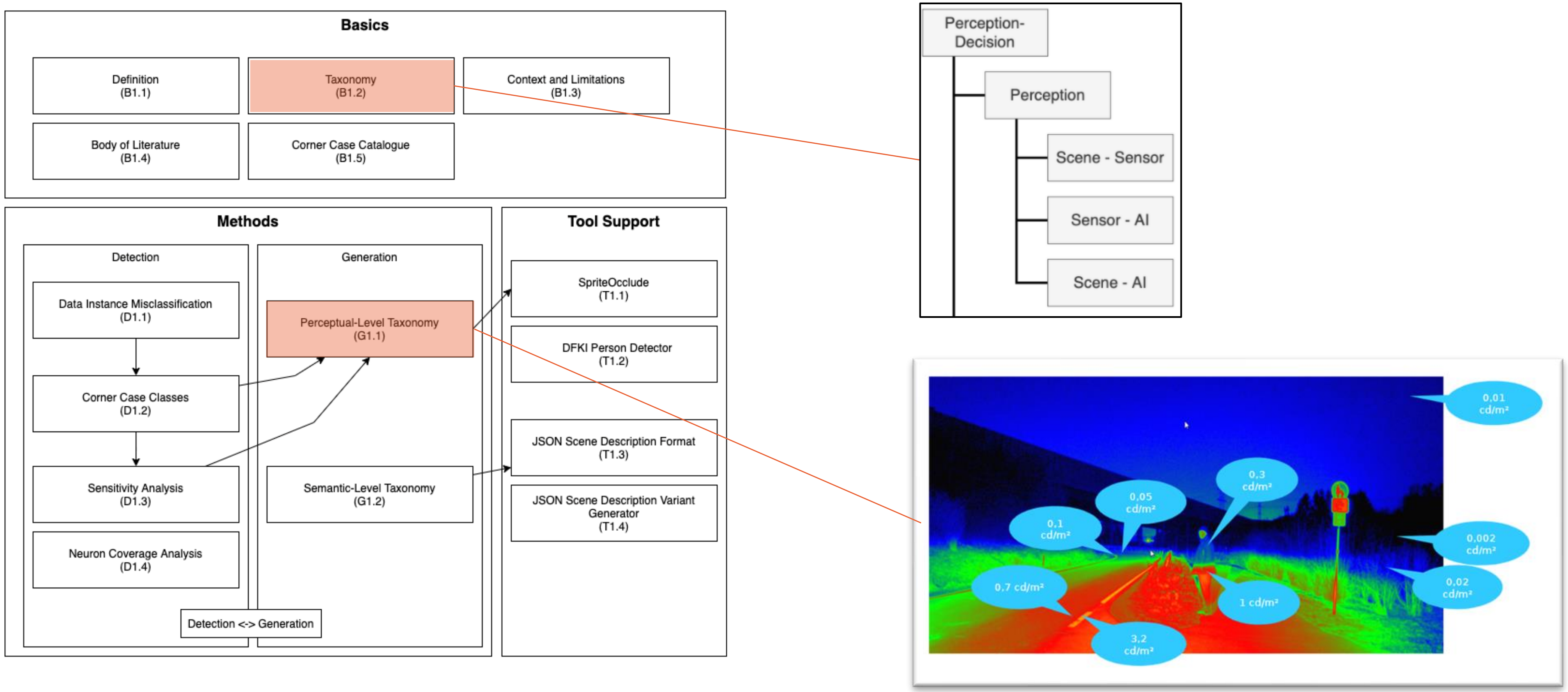


Frames + Metadata



Supporting: data specification format (JSON)

Method framework for systematic elicitation and analysis of corner cases



Intermediate key findings in data generation

- The generation of a large data amounts in high quality with rich annotations and meta data is possible.
- High one-time development effort is required to meet typical requirements.

- We are looking forward to using the new toolchain features and to the further learnings from data quality analysis, corner case analysis and safeguarding in general.

KI ABSICHERUNG

Safe AI for Automated Driving

Dr. Thomas Stauner, BMW AG

Thomas.Stauner@bmw.de

KI Absicherung ist ein Projekt der KI Familie
und wurde aus der VDA Leitinitiative autonomes
und vernetztes Fahren heraus entwickelt.

KI FAMILIE

Gefördert durch:

Bundesministerium
für Wirtschaft
und Energie

aufgrund eines Beschlusses
des Deutschen Bundestages

www.ki-absicherung.vdali.de @KI_Familie KI Familie

Appendix: Sensor abstraction. Domain transfer example.

- Exploration of domain adaptation techniques
  - Adaptation of data from one domain to another
    - Idea: Generate scenes on higher level of abstraction (Scene Graphs)
    - 2 Generator Networks infere semantic and spatial layouts from scene graph description
    - 1 Adversarial Network renders “real world” frame from layouts