KI DATA TOOLING
The Data Kit for Automotive AI

KI Data Tooling Final Event | 05/06 December 2023

Data centric AI Developer Journey in the KI Data Tooling Framework

Thomas Stone, BMW | Jasmin Breitenstein, TU BS |Evren Ermiş, Continental | Jakob Kirner, BIT TS

AI Developer Journey

- Data generation & usage framework developed.
- How to navigate through the process as a developer?
  - Example from the project:
    - „Honey, I partially occluded the kids“
    - … and now car does not see them anymore!
- Let‘s **follow our Dev** with a common problem:
  - **Occluded pedestrians!**

1

Occluded pedestrians are hard

Amodal Perception: From a Magic Trick to Automated Driving

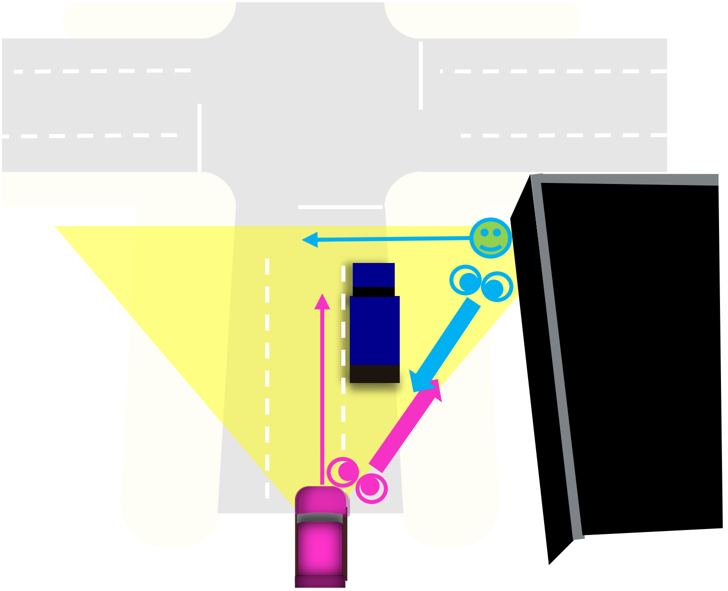
- The parrot does not understand that his owner just vanished behind the wall
- … however must of us humans are able to grasp **what is happening** in this magic trick
- Amodal perception is the ability to hallucinate the full shape of objects behind (partial) occlusions
- **Humans are very good at this**
- Much like the parrot, our **perception functions** in automated driving are **not able to understand occlusions**
- Perception methods often fail when confronted with occlusions

[Video source: youtube, LADbibleExtra, “Parrots gets confused by What the Fluff”]

Amodal Perception: Importance in Automated Driving

- Amodal perception is the ability to hallucinate the full shape of (partially) occluded objects
- Automated driving: perceiving occludees behind occlusions is **crucial for safe environment perception**
- **Humans are good at this, perception methods in general not**

Example scenario for occlusion in automated driving:

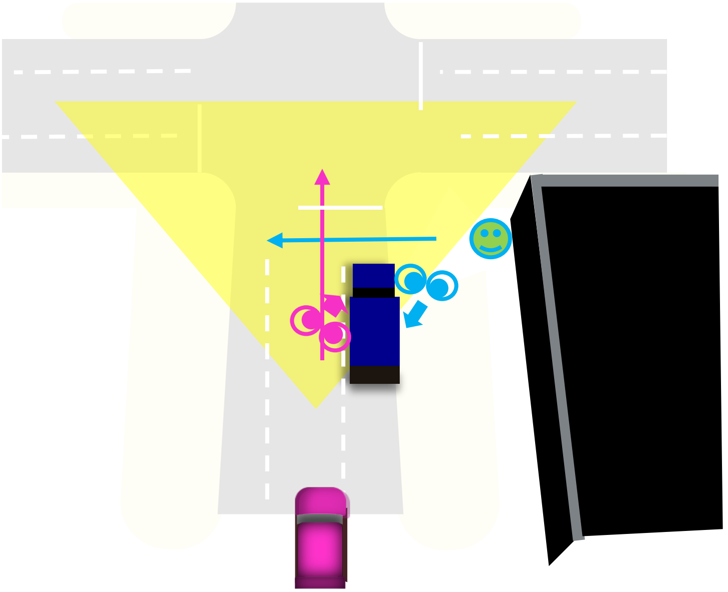


A truck is parked on the side of the street.
Pedestrian is walking towards the street.
Ego-vehicle is driving towards the intersection.
Pedestrian and ego-vehicle see each other.

Amodal Perception: Importance in Automated Driving

- Amodal perception is the ability to hallucinate the full shape of (partially) occluded objects
- Automated driving: perceiving occludees behind occlusions is crucial for safe environment perception
- Humans are good at this, perception methods in general not

Example scenario for occlusion in automated driving:

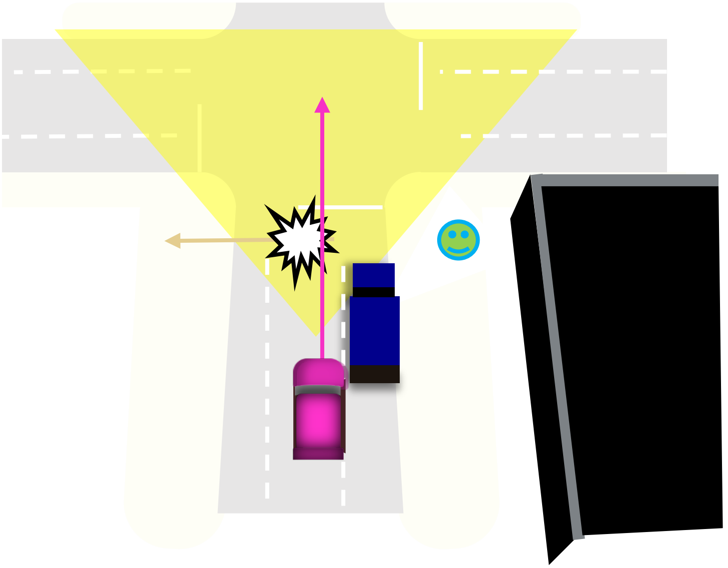


A truck is parked on the side of the street.
Pedestrian is walking towards the street.
Ego-vehicle is driving towards the intersection.
Ego-vehicle is now next to the truck.
Pedestrian and ego-vehicle do not see each other.

Amodal Perception: Importance in Automated Driving

- Amodal perception is the ability to hallucinate the full shape of (partially) occluded objects
- Automated driving: perceiving occludees behind occlusions is **crucial for safe environment perception**
- **Humans are good at this, perception methods in general not**

Example scenario for occlusion in automated driving:



A truck is parked on the side of the street.
Pedestrian is walking onto the street.
Ego-vehicle is driving towards the intersection.
Ego-vehicle and pedestrian reach the end of the truck.
Pedestrian and ego-vehicle see each other again but it is too late.

Conventional Perception Methods Fail at Occlusions

- In contrast to humans, conventional perception methods cannot perceive or understand occlusions

Testing accuracy of object classification methods under extreme occlusion on the VehicleOcclusion dataset

[Zhu et al., 2019]

Humans/Methods	w/o occlusion	w/ occlusion
Humans	-	**93.3%**
AlexNet	89.8%	50.0%
ResNet	90.1%	54.0%
VGG16	**94.7%**	**62.6%**

example images

- Worse perception of occluded objects results in corner cases
- Other types of corner cases can form behind occlusions

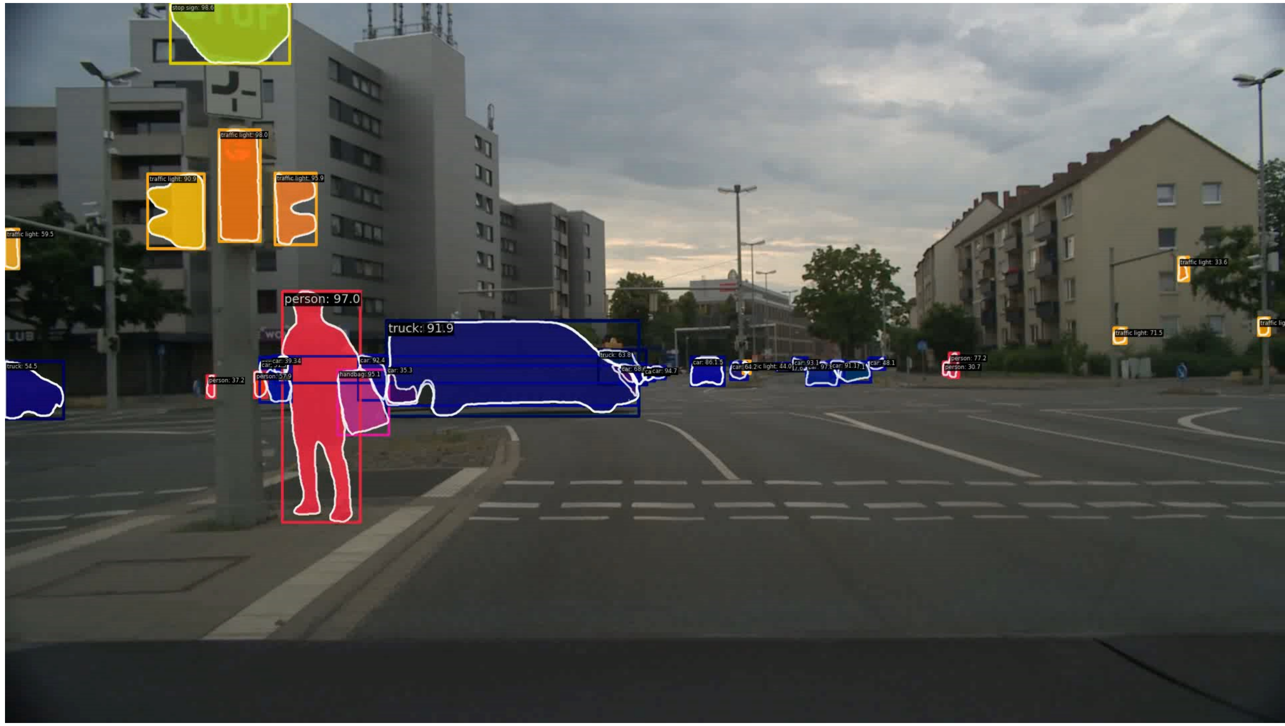
→ Amodal methods learn to anticipate occluded objects

Conventional Perception Methods Fail at Occlusions

Intuition (qualitative) for the failure of segmentation methods on KI-DT data

- Instance segmentation using Mask2former [Cheng, 2022] on KI-DT data
- Clemens is close to the ego-vehicle
- He is **visible in the beginning** and the **segmentation works**
- Clemens is walking behind the pole
- He **can no longer be detected** by Mask2former because he is **too occluded**
- Some body parts of Clemens are segmented, but seem like **people in the background** → critical for environment perception and understanding

→ We need **amodal perception** for **safe perception** of occluded objects



2

Let‘s get more data!
You know, like always?

More sample data where?

- Just get **more samples** of problematic corner cases!
- How can we **find** the right data?
    - **Context**
    - **AI-based search**

Journey steps from data to product

- **Clear representation** of the datasets content supports discoverability of the dataset
- **Clear specification** of the dataset supports usability of the dataset
- **Clear access management** of the dataset by the data owner
- How to **support the domain expert** to become a data owner?

Data Mesh

- Decentralized approach for a data architecture
- Covers the key aspects of a data landscape with many contributors and consumers
- It‘s core concept of a **data product** owned with **domain ownership** enables a new sozio-technological approach to today‘s highly diverse data requirements

Principles of the Data Mesh Concept

Domain ownership

Data owned by its domain centered product team. Own your data, own your product.

Data as a product

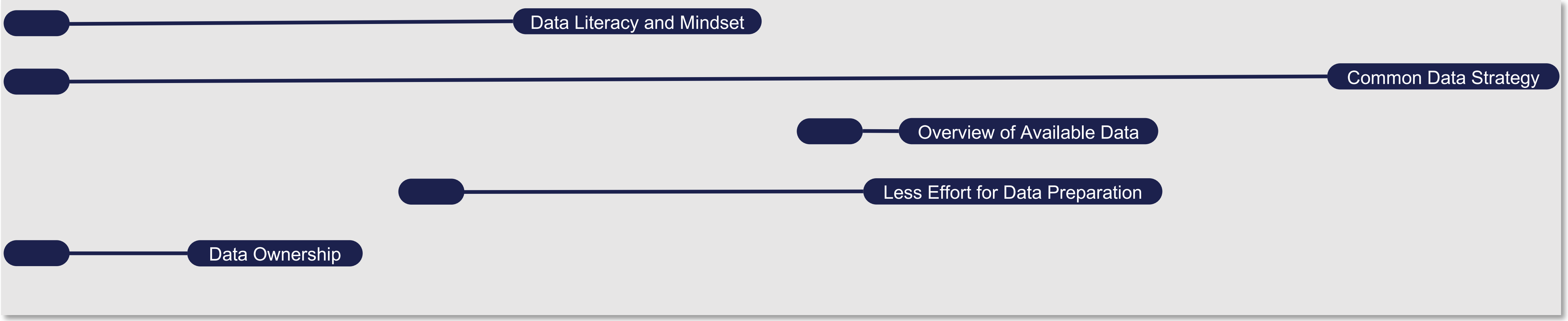
Think beyond data as an asset. Design thinking for data.

Self-serve data platforms

Reduce effort and technological complexity to support interaction between producer and consumer.

Federated computational governance

Decentralization and domain self-sovereignty, interoperability through global standardization



KI DATA TOOLING
The Data Kit for Automotive AI

Clear Representation

Clear Representation

- Context is metadata about a recording in the KI-DT dataset
- Machine-readable (JSON)
- Can describe the entire recording or one of the frames

Statistics → 1 million frames with 200 million context attributes

Types of context

Data source	Attributes
CAN bus data	Location, speed, acceleration, … of the ego vehicle
Map database (e.g. OpenStreetMap)	Nearby objects, road attributes
Local time	Sunlight angles, holidays and special events
Stationary sensors	Traffic intensity, criticality, weather station data
Camera images	Weather and lighting, maneuvers, caption
Lidar data	Precipitation, spray
Labels or synthetic data generation	Assets, road users

Context generation process

- Sequential application on shared plattform
- Uses different tools and technologies to generate a single consistent result

`context-ident-structured-data`

`context-ident-weather`

`context-ident-sun-position`

…

`validator`

`context.json`

Clear Specification

Clear Specification

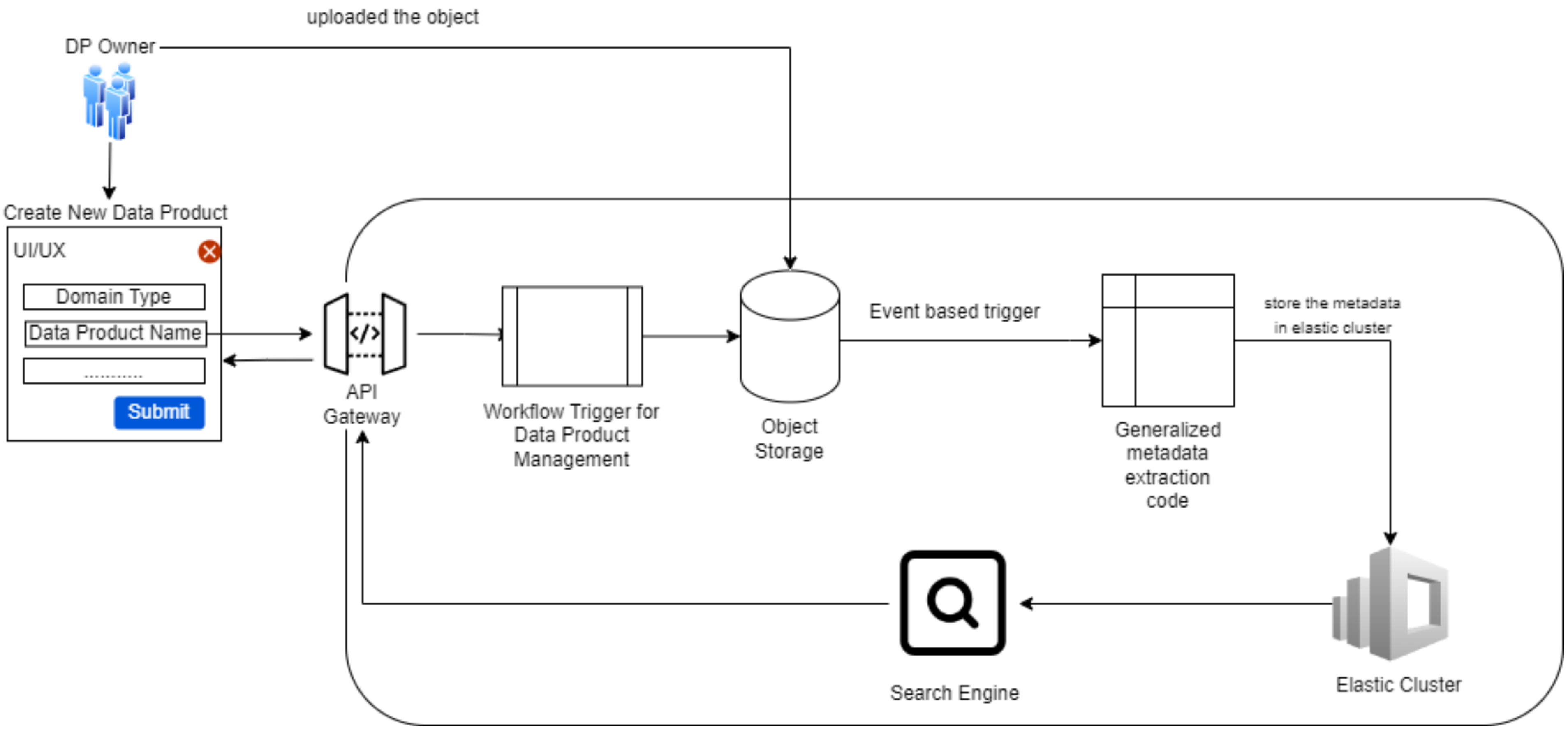
- Based on KI Absicherung SynPeDS
- Extended by real and augmented data
- More information in KI Data Tooling Data Spec

Table of Content - KI Data Tooling Data Spec

Support the Domain Expert

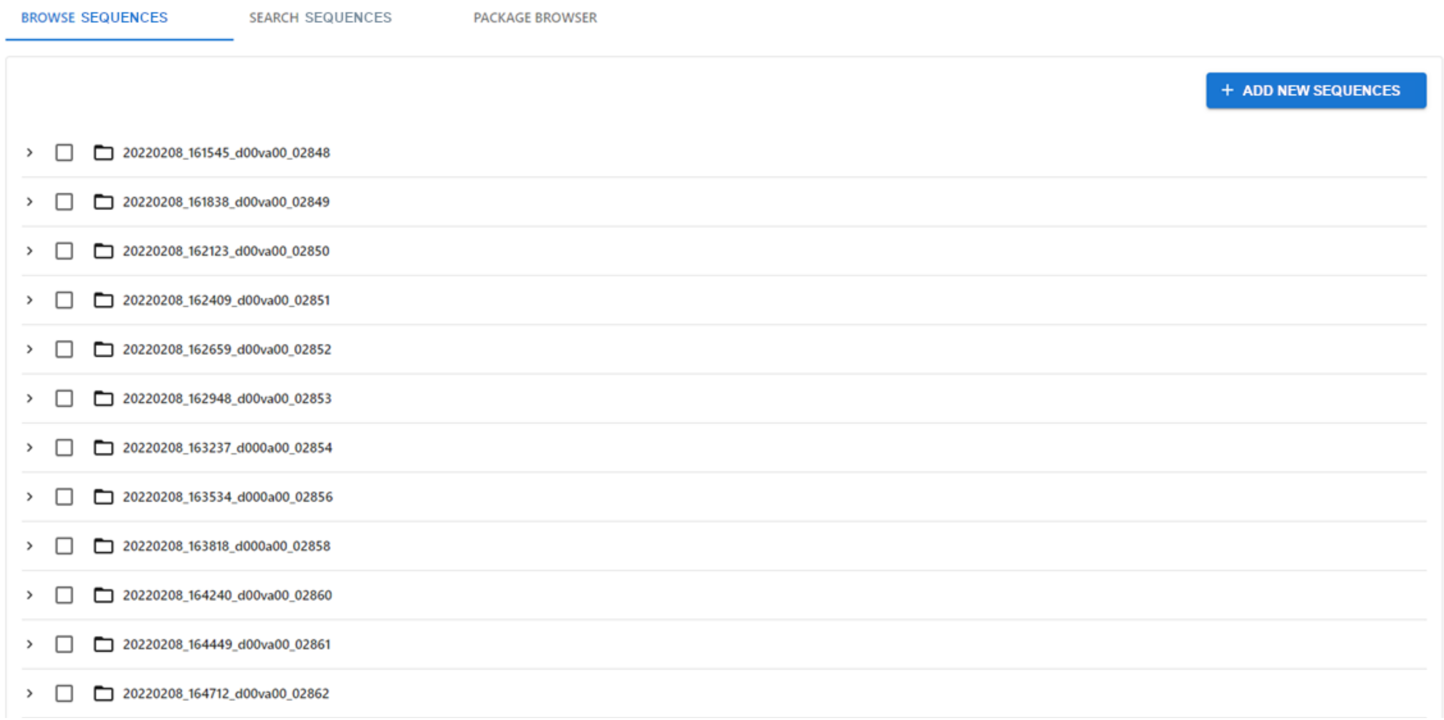
KI DATA TOOLING
The Data Kit for Automotive AI

How to support the domain expert to become a data owner?

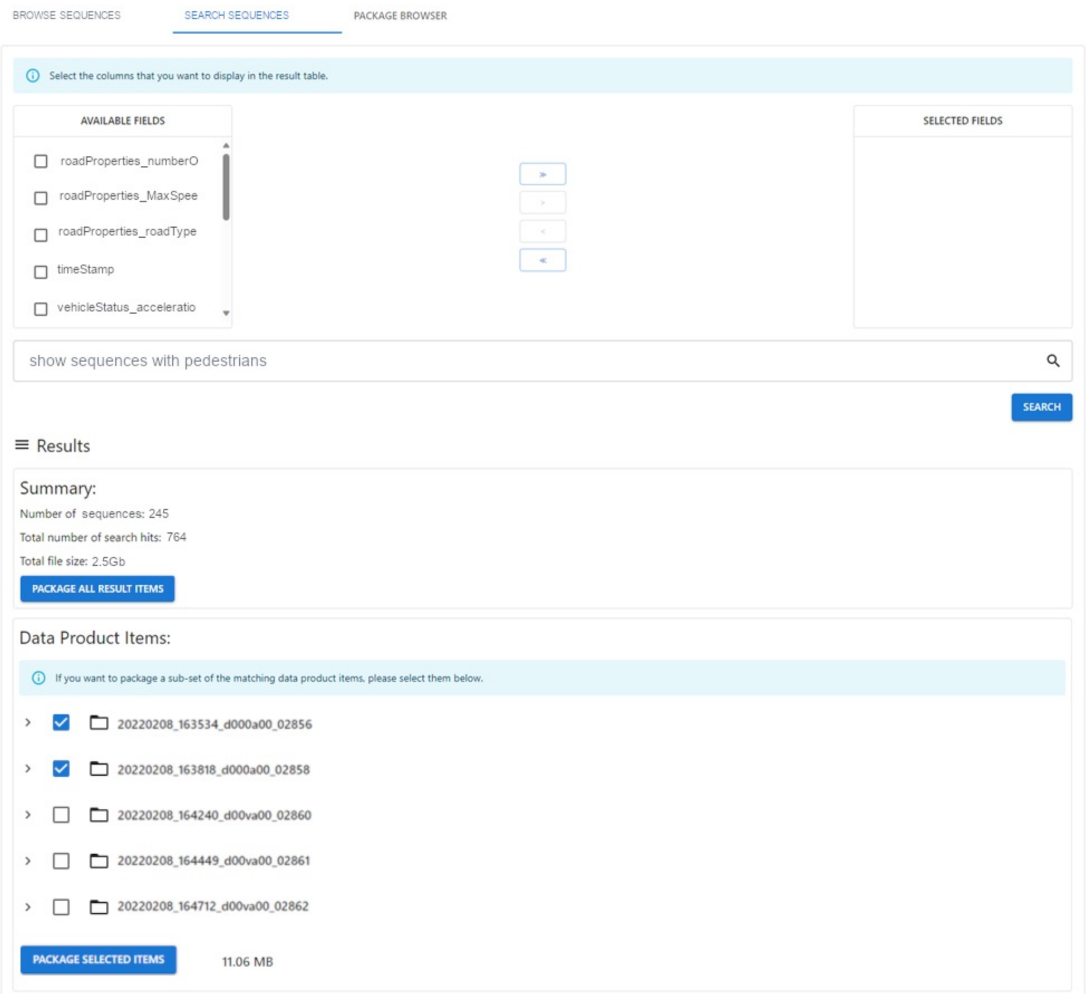


Discoverability (1)

Content of the data product



Discovering data in catalogue



Provisioning the requested data

BROWSE SEQUENCES SEARCH SEQUENCES PACKAGE BROWSER

Packaging Jobs

FILTERS

Job Name	Creator	Creation Date	Search Query	Number of items	Total Size	Packaging Status	Actions
myjob	Jan	2.11.2023, 09:25:41	give me all data for SDG	6	500.00 MB	Completed	
myjob1	Jan	24.11.2023, 10:04:08	headposition_x > 0	2	213.90 MB	Pending	
myjob2	Jan	2.11.2023, 09:25:41	give me data where there i...	120	7.72 GB	Completed	

Rows per page: 100 1–3 of 3

Clear access management

User Management

Manage the users of your data product.

Subscribers

Subscribers can use all of the data products features in a read only mode.

Consumer	Request Date	Changed Date	Actions
Evren	31.10.2023, 11:11:13	31.10.2023, 11:11:24	

1–1 of 1

1 Pending requests

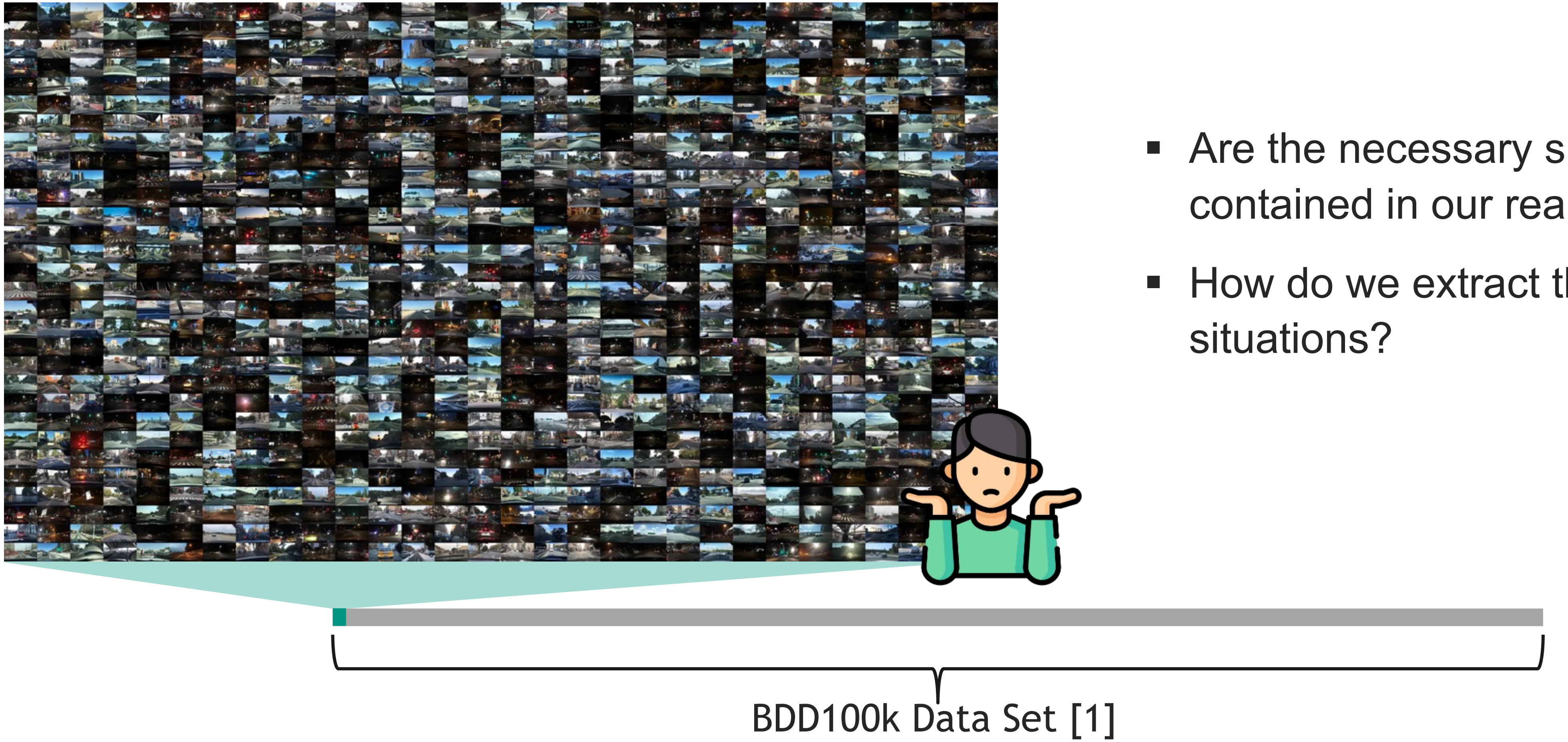
Consumer	Request Date	Request Type	Actions
Tom	24.11.2023, 10...	subscribe	

1–1 of 1

Discoverability (2)

How to find context which is not labled?

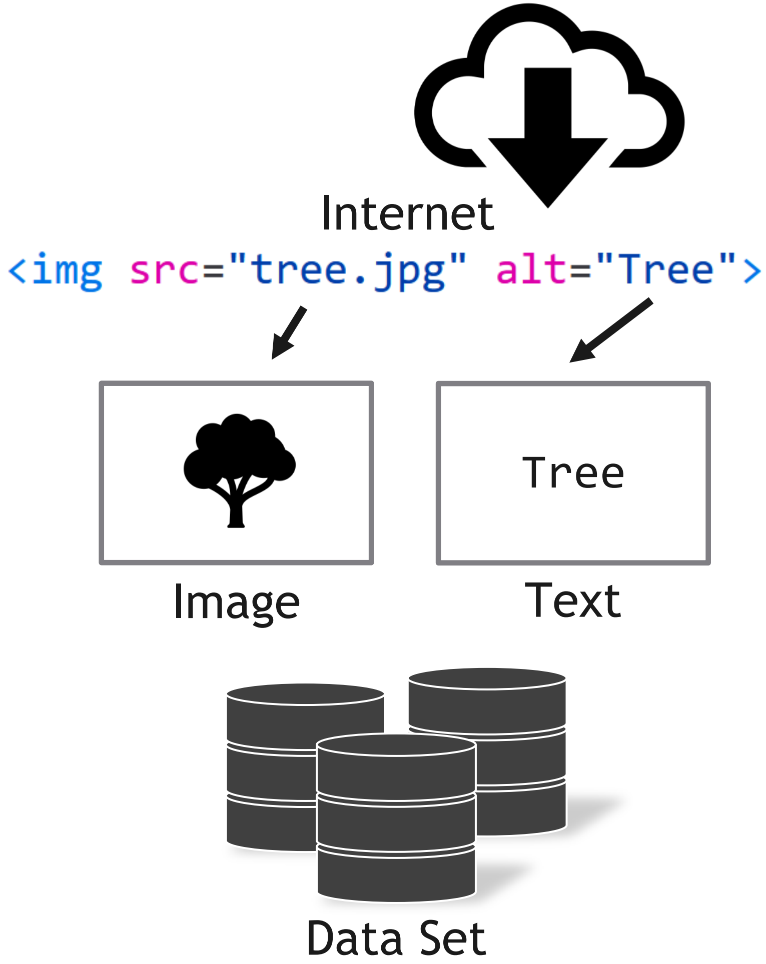
Manual search in the data:

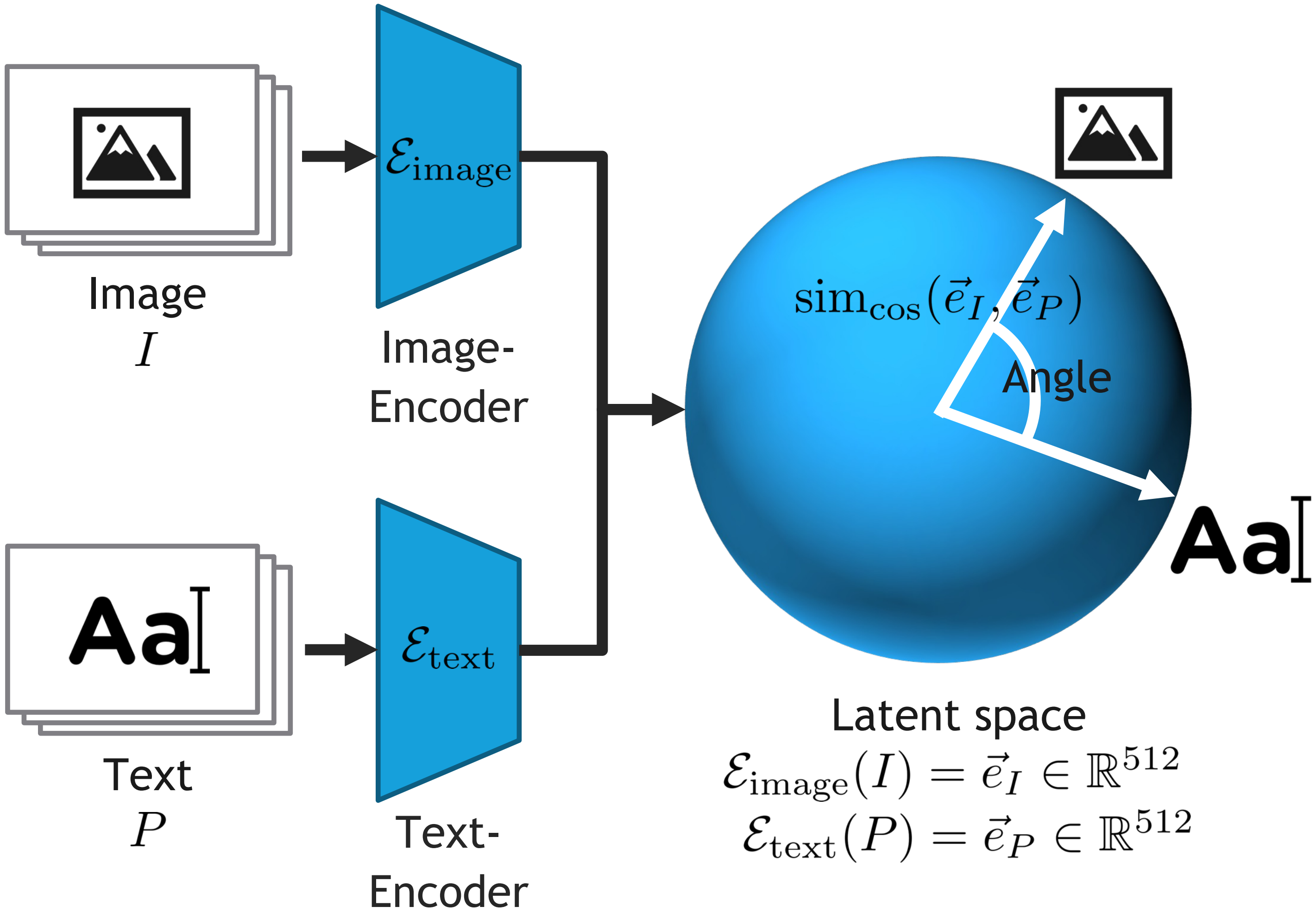


- Are the necessary situations contained in our real data?
- How do we extract the relevant situations?

[1] F. Yu u. a., „BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning“, arXiv:1805.04687 [cs], Apr. 2020.
Icon made by Freepik from Flaticon.com

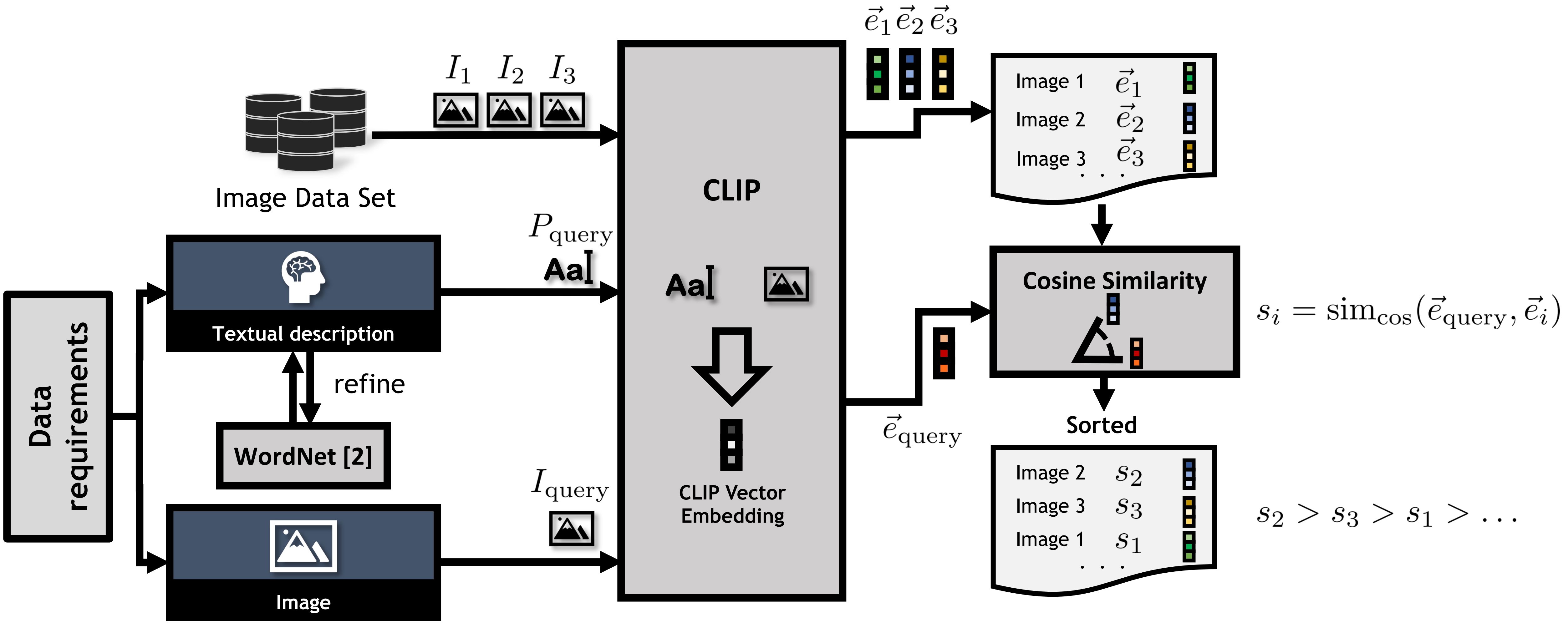
Generic Data Search
Foundation: CLIP [1]





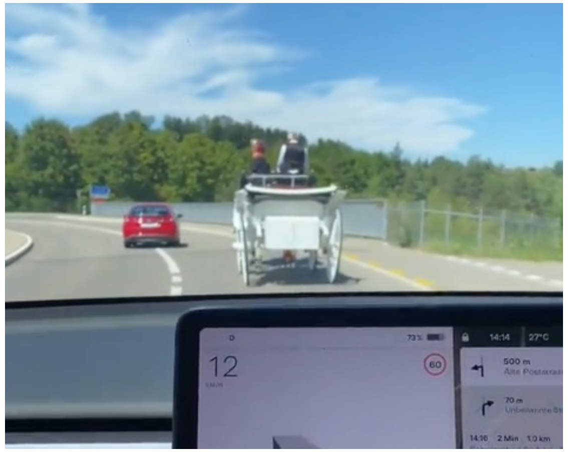
[1] A. Radford u. a., „Learning Transferable Visual Models From Natural Language Supervision“, S. 16, 2021.

Generic Data Search [1]



[1] P. Rigoll et al. „Focus on the Challenges: Analysis of a User-friendly Data Search Approach with CLIP in the Automotive Domain“. arXiv, 21. April 2023. [Online].
[2] G. A. Miller, „WordNet: a lexical database for English“, Commun. ACM, Bd. 38, Nr. 11, S. 39-41, Nov. 1995.

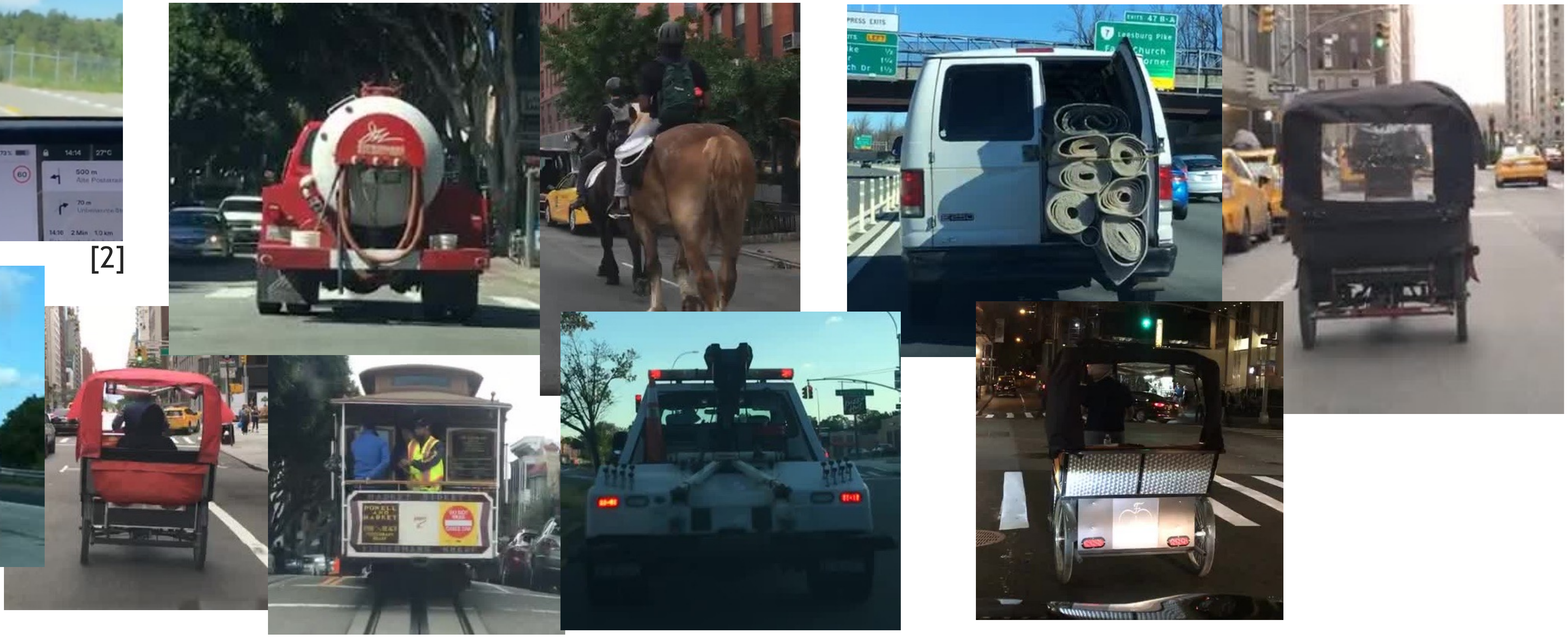
Generic Data Search



[2]

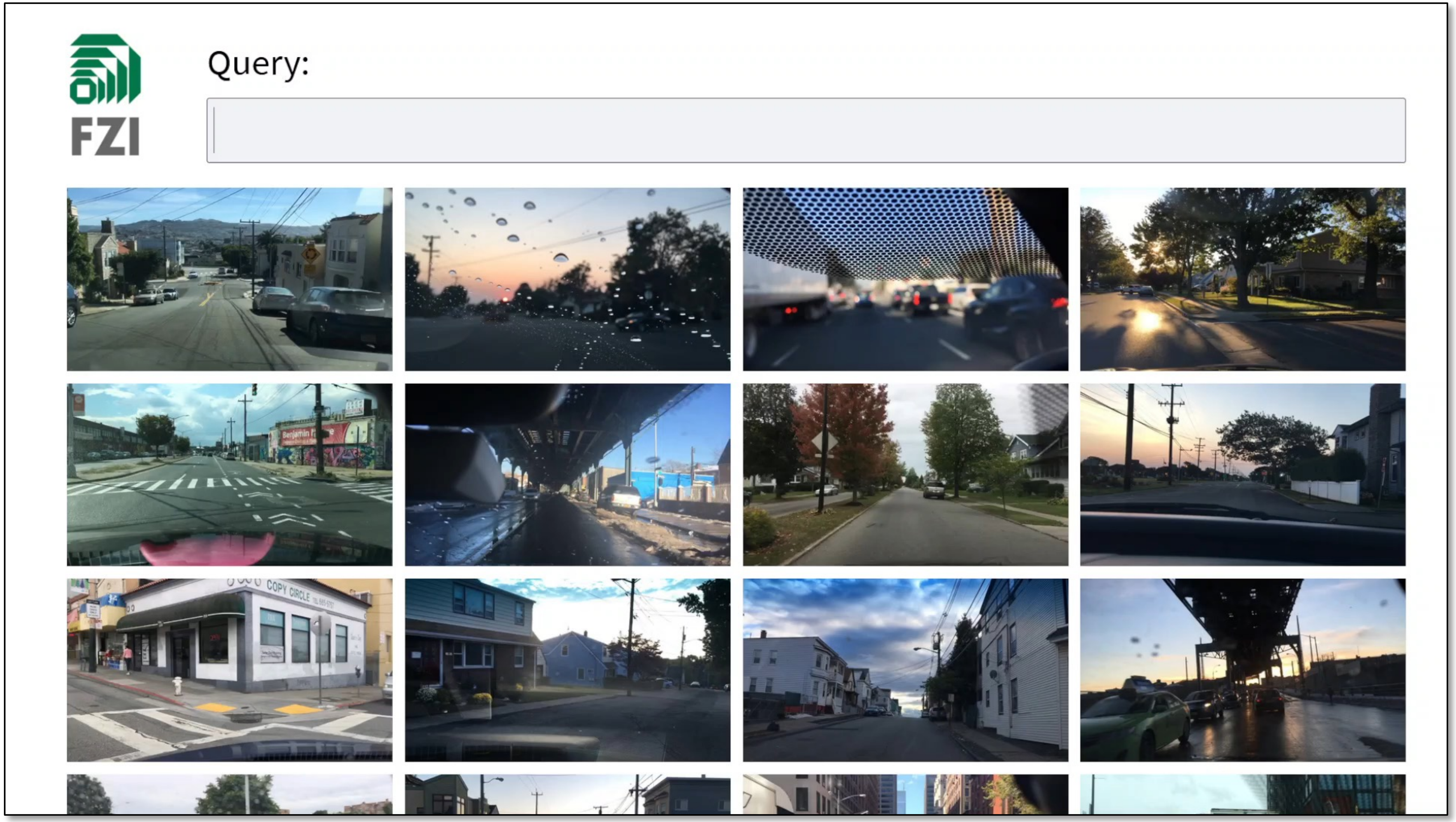
carriage → autorickshaw, tuktuk, special vehicle, horses

Related words (WordNet [1])



[1] G. A. Miller, „WordNet: a lexical database for English“, Commun. ACM, Bd. 38, Nr. 11, S. 39-41, Nov. 1995.
[2] _realrusty: „TeslaVSPferdekutsche“. Abgerufen 22.08.2022. https://www.tiktok.com/@_realrusty/video/7131351993859329285
Other images: F. Yu u. a., „BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning“, arXiv:1805.04687 [cs], Apr. 2020.

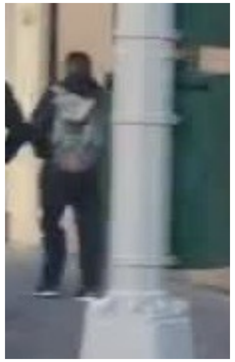
Generic Data Search



Images: F. Yu u. a., „BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning“, arXiv:1805.04687 [cs], Apr. 2020.

Generic Data Search

Prompt: `image of pedestrian occluded by a post`



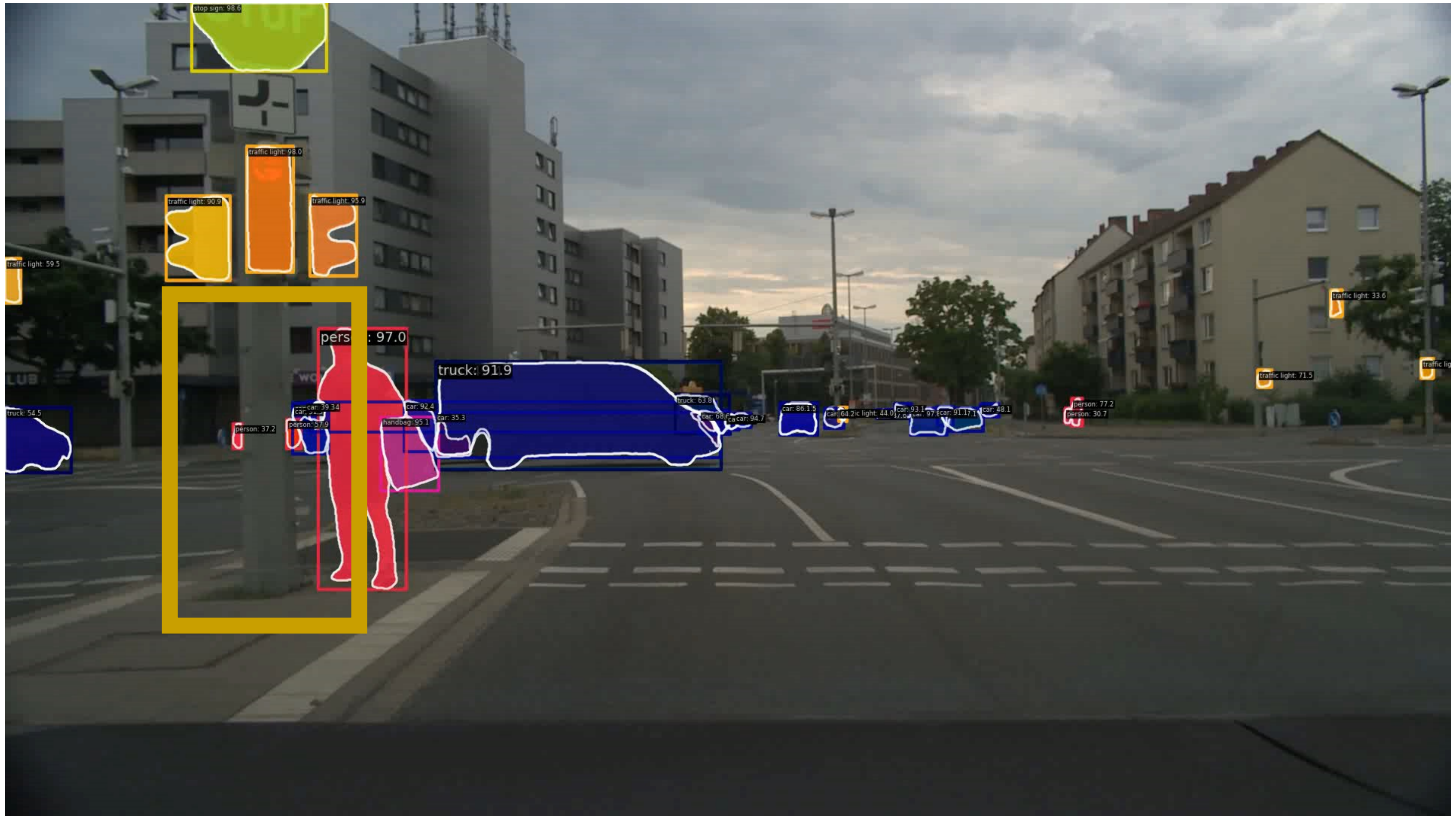
Images: F. Yu u. a., „BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning“, arXiv:1805.04687 [cs], Apr. 2020.

3

OK, now what do we want?
Improvise. Adapt. Overcome.

Conventional Perception Methods Fail at Occlusions

Let’s look again at this example for the failure of segmentation methods on KI-DT data



[Cheng et al., “Masked-attention Mask Transformer for Universal Image Segmentation”, in Proc. of CVPR 2022]

- Instance segmentation using Mask2former[Cheng, 2022] on KI-DT data
- Ideally we would want not only a ground truth for what is visible, but also a ground truth mask for the occluded parts
- With this ground truth we could train perception methods that can perceive occlusions

Requirements for Amodal Segmentation Data

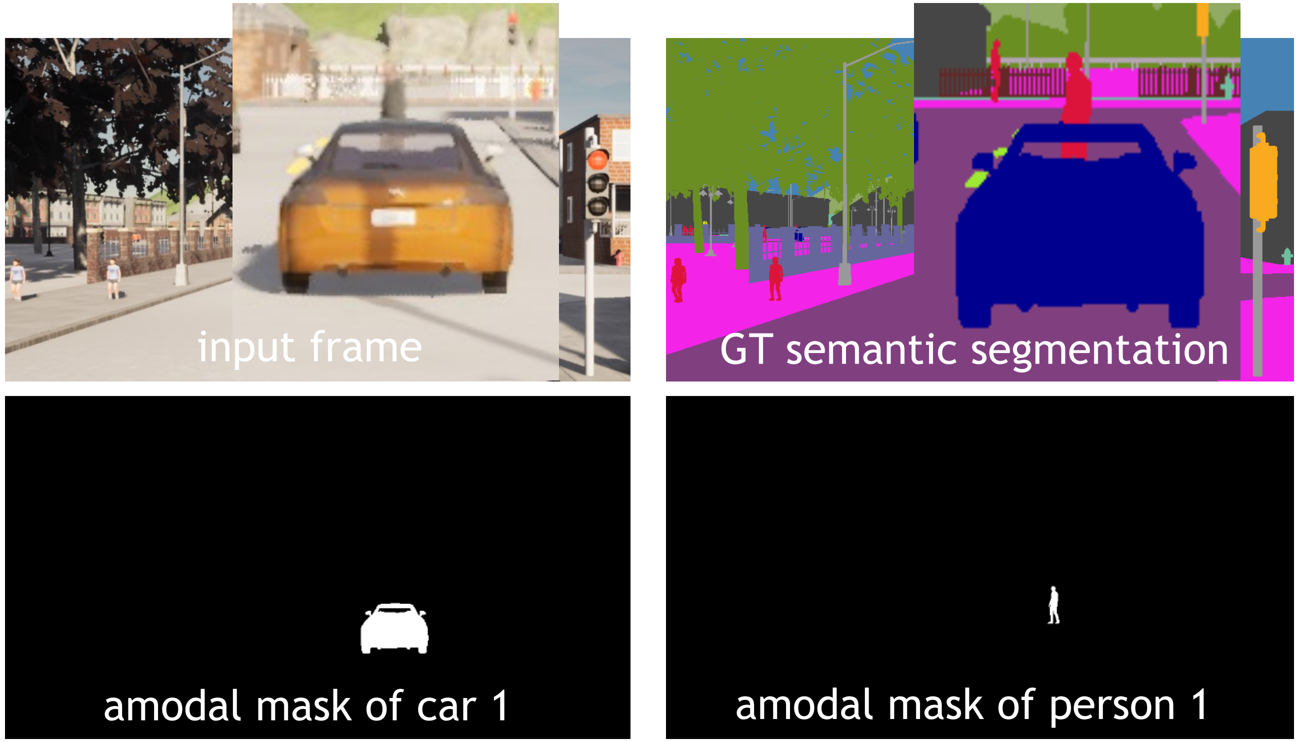
1. We need occluded objects
2. We need the standard visible segmentation
   - One class label per pixel that is visible in image
3. We need the ground truth of occluded objects
   - One class label per pixel which is occluded in the image

For amodal segmentation on single images, we can generate synthetic amodal data using copy-paste methods

Requirements for Amodal Segmentation Data

- On videos, requirements change as we need to **identify specific instances across all frames of a video**
- Per frame and per instance we need both amodal and visible instance mask

Example frame from the CARLA simulator:



Annotated amodal masks shown for two instances: car 1 and person 1
→ On videos **differentiating between different instances of the same class** is important due to different patterns of movements

Requirements for Amodal Segmentation Data

Training of Perception Functions

- Additional requirements from training, validation and evaluation of perception functions
- Data amount needed derived from existing datasets to train segmentation methods on images and videos

Dataset statistics for amodal image dataset:

	Cityscapes	Amodal Cityscapes
#images (train)	2975	2900
#images (val)	500	75
#images (test)	1525	500

→ Dataset with amodal ground truth should follow their visible counterpart datasets in terms of size and other parameters

→ So far no automotive amodal video dataset has been published

Dataset statistics for amodal video dataset:

	BDD MOTS	OVIS	SAIL-VOS
Synthetic/Real	Real	Real	Synthetic
Amodal GT	No	No	Yes
#videos	186	901	201
#frames	37,220	69,035	111,654
Average length	40s	12.77s	69.38s
defined data splits	No	Yes	Yes
#instances	22,963	5,223	1,896,296

[J. Breitenstein, T. Fingscheidt, "Amodal Cityscapes: A New Dataset, its Generation, and an Amodal Semantic Segmentation Challenge Baseline," In Proc. IV, 2022.]

Requirements for Amodal Segmentation Data
Towards a Challenge for Amodal Segmentation

This we could publish

On Images:

Training: Amodal Cityscapes training data
Validation: Amodal Cityscapes validation data
Evaluation: Amodal Cityscapes test data

Evaluation Metrics:
common metric for semantic segmentation mIoU and $mIoU_{inv}$ (mIoU on occluded areas)

Check out our paper !

[J. Breitenstein, T. Fingscheidt, "Amodal Cityscapes: A New Dataset, its Generation, and an Amodal Semantic Segmentation Challenge Baseline," In Proc. IV, 2022.]

On Videos:

Problem: So far the only dataset with amodal ground truth on videos is SAIL-VOS but it is non-automotive

Challenge KI Data Tooling:
Generation of suitable automotive data for training and evaluation, baseline published [Breitenstein et al., 2023]

Evaluation Metrics (derived from SAIL-VOS challenge):
Common metric for video instance segmentation: mAP and mAP_{50} and mAP_{50} on partially and heavily occluded objects

[J. Breitenstein, K. Jin, A. Hakiri, M. Klingner, T. Fingscheidt, "End-to-end Amodal Video Instance Segmentation," In Proc. BMVC-Workshops, 2023.]

4

Push-button gene*ray*tion?

Ticketing the project toolchain

#059 AP2.3 Synthetic Data for Amodal Segmentation

Created by Jasmin Breitenstein, last modified on Feb 24, 2023

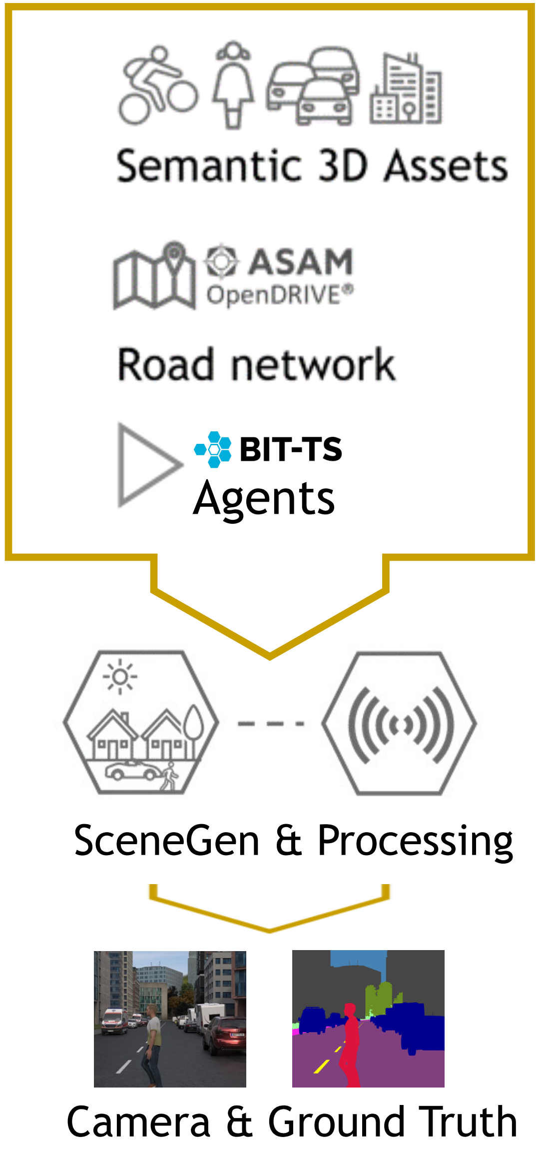
Data source (car & infrastructure)

- [ ] Valeo
- [ ] Bosch
- [ ] ZF/FZI
- [x] BIT
- [ ] junction Aschaffenburg
- [ ] junction Braunschweig

Requirements for synthetic data:

- Urban scene, comparable to the *Cityscapes* dataset
- Occluded VRUs in relevant positions, proximity to ego vehicle
- 50 scenarios with 50 frames and 4-6 variations each
- 30-10-10 split for training, testing and validation per scenario
- → max **15.000 frames**

Toolchain view



Ground Truth Generation with OspRay

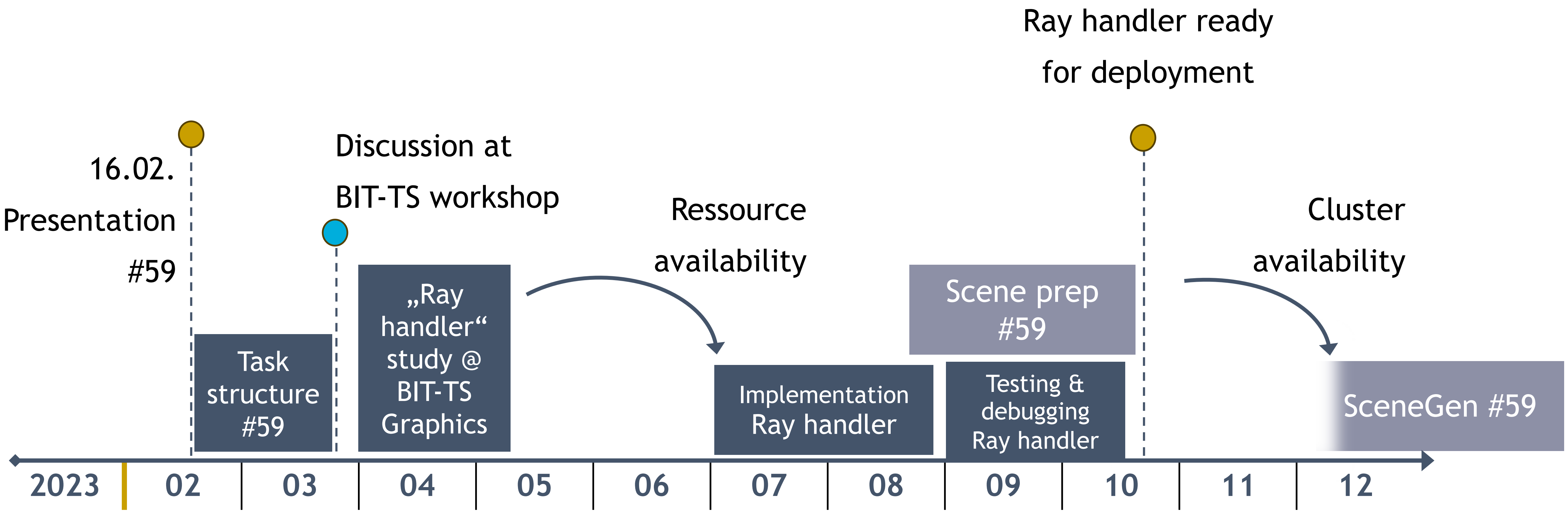
Ground Truth is generated with metadata plugin

- Metadata is exported in EXR layers and JSON files
  - Bounding boxes
  - Semantic segmentation
  - Depths
  - Instance segmentation

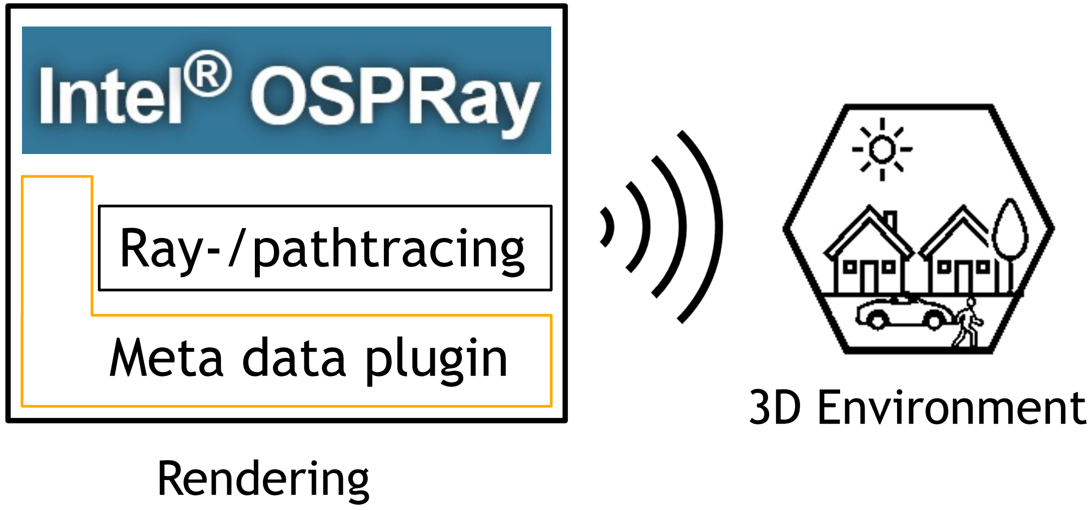
Introducing the ray handler

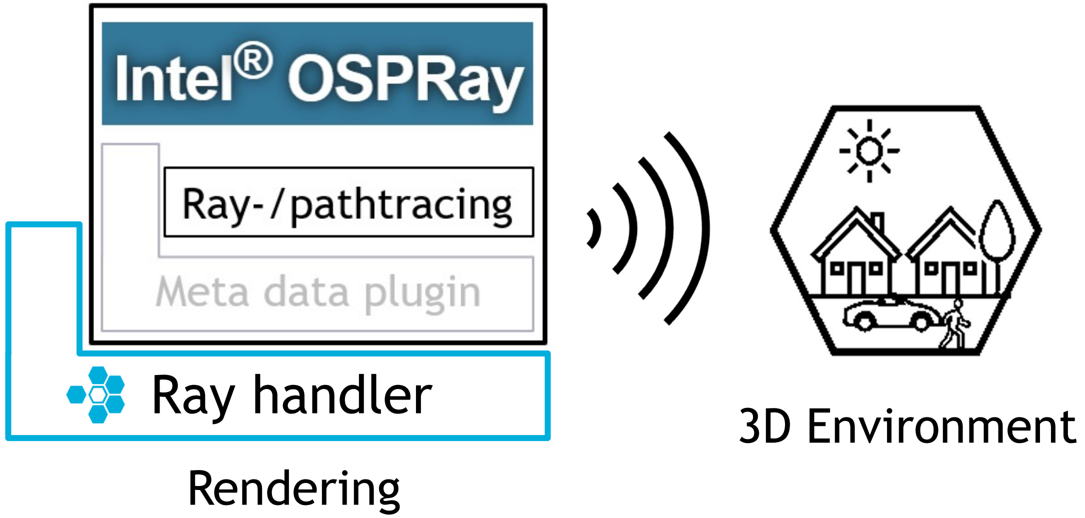
/imagine prompt A photorealistic picture of a ray handler handling rays

Ground truth overhaul



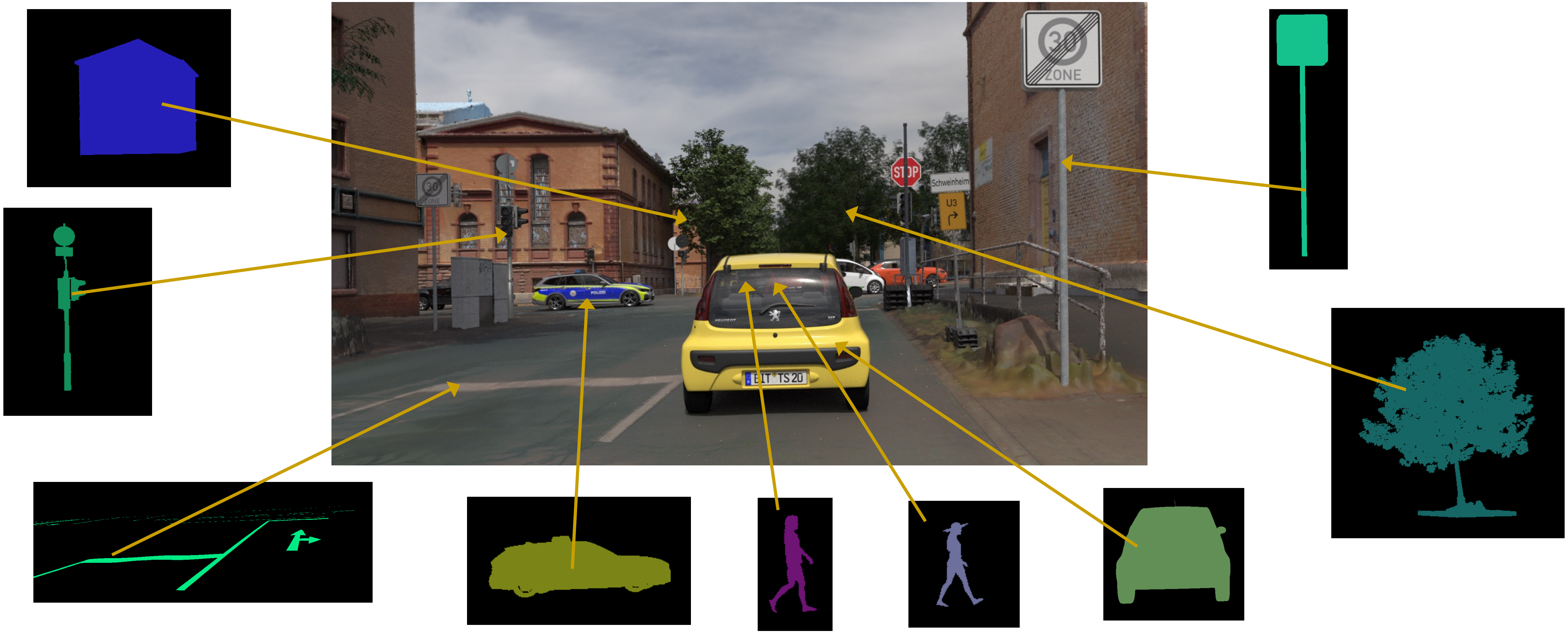
Ground truth overhaul





Occlusion representation before #59	Occlusion representation after #59
- Occlusion roughly estimated by asset geometry - Multiple overlaps not representable with current OSPRay meta data plugin	- Occlusion calculated by asset geometry - Pixel accurate occlusion annotation - New GT formats like order of objects

Amodal Segmentation & Masks



Shown subset of in total **77 amodal masks** in this particular frame, 2 occluded pedestrians

5

Now generate and archive!

Data Curation

- After data generation ➔ data ingest + curation

6

Are we better now?

Training with Amodal Data

Results of our Amodal Segmentation Methods on Images

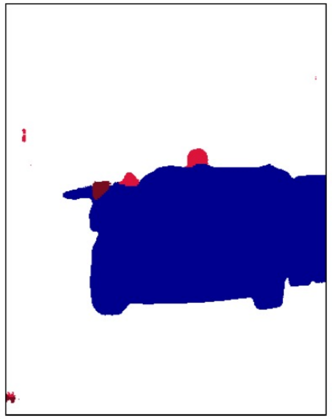
Performance of standard visible semantic segmentation, only amodal semantic segmentation and joint amodal and visible semantic segmentation of the Amodal Cityscapes test data and the Cityscapes validation data. Best results in **bold**.

Method	Amodal Cityscapes test data		Cityscapes validation data	
	mIoU	$\mathrm{mIoU}^{\mathrm{inv}}$	mIoU	$\mathrm{mIoU}^{\mathrm{inv}}$
Standard visible segm.	62.99%	5.00%	67.21%	cannot be calculated due to missing ground truth
Only amodal segm.	20.16%	36.48%	21.00%	
Joint amodal and visible segm. (ours)	**63.32%**	**43.32%**	**68.35%**	

Joint training improves both visible and amodal performance

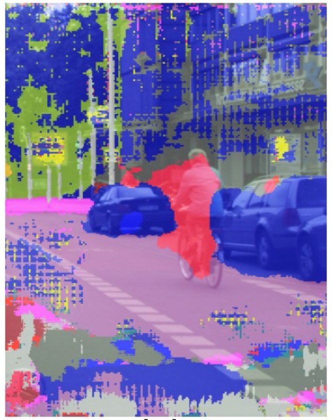
Qualitative results of the joint method:

input image

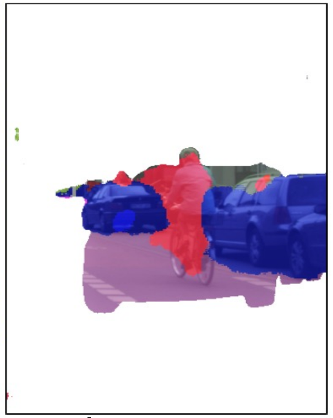


mobile objects

sem. segm.



amodal segm.

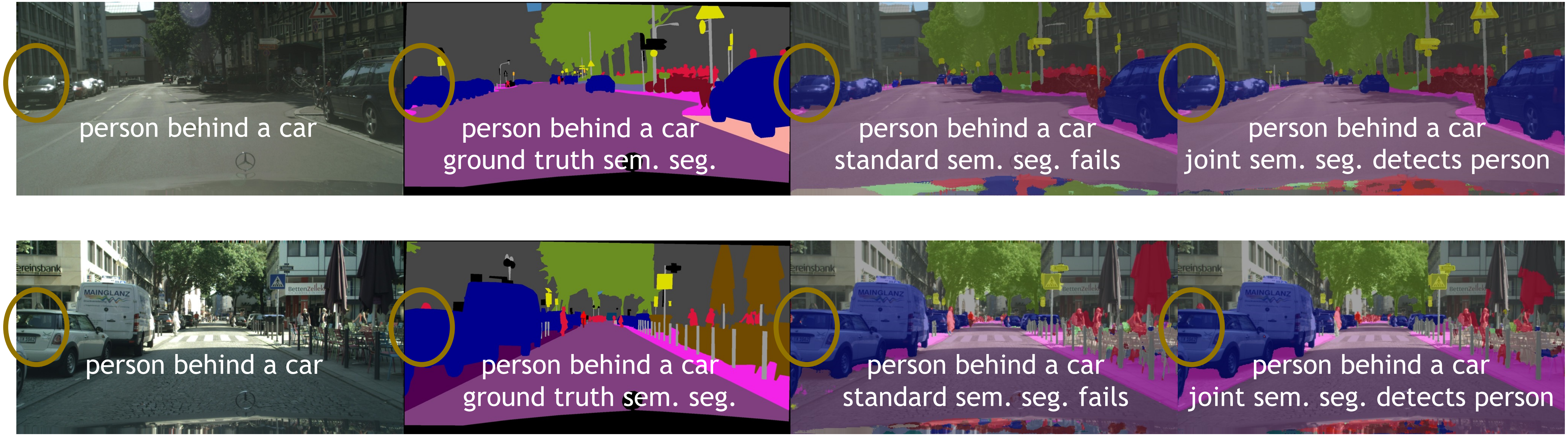


relevant parts

- Mobile objects = all moving predicted objects: Person, rider, car, truck, motorcycle, bicycle, bus, train
- Relevant part: Insertion of amodal prediction into the predicted mobile objects
- Occluded person behind the car is anticipated by joint training

Training with Amodal Data

Qualitative Observations in Pedestrian Detection



We observe during evaluation a better segmentation performance of pedestrians for the joint visible and amodal semantic segmentation and, qualitatively, better segmentation of occluded pedestrians

Training with Amodal Data
Amodal Segmentation Methods on Videos

- So far, only image-based methods for amodal segmentation exist
- BUT: Additional temporal context allows us to better treat heavy occlusions
- To our knowledge SAIL-VOS is the only dataset with amodal ground truth annotations on video level

→ We investigate end-to-end amodal video instance segmentation on the SAIL-VOS data

Example validation video with amodal masks, bounding boxes and class labels provided by our end-to-end amodal video instance segmentation (VATRACK) baseline:

Re-identified instances are visualized in the same color

Interested in our methods?

Deep Dive #4
Corner Case

→ We see that VATRACK is able to provide enough temporal context to follow the occluded person (grey) as well as the visible person (blue)

→ For automated driving, we require synthetic data with amodal ground truth to show this application for critical occlusions in automated driving

7

Now all good?
What‘s to come?

Mixed Training
Identification and Filling of Gaps in Data

Goal:

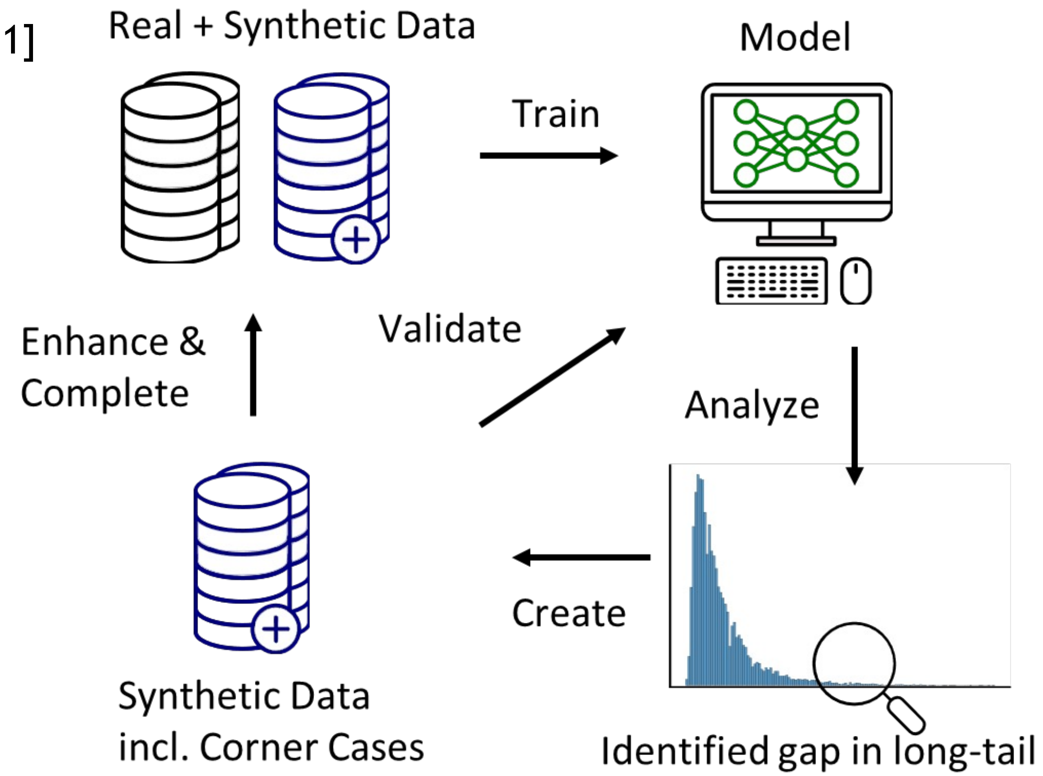
Integrate synthetic data into data-driven engineering Process [1]

Paradigm Shift:
Generate the data the model really needs!

Note: This is possible with **synthetic data**, as we do have **control over the data** generating process.

Research Questions:

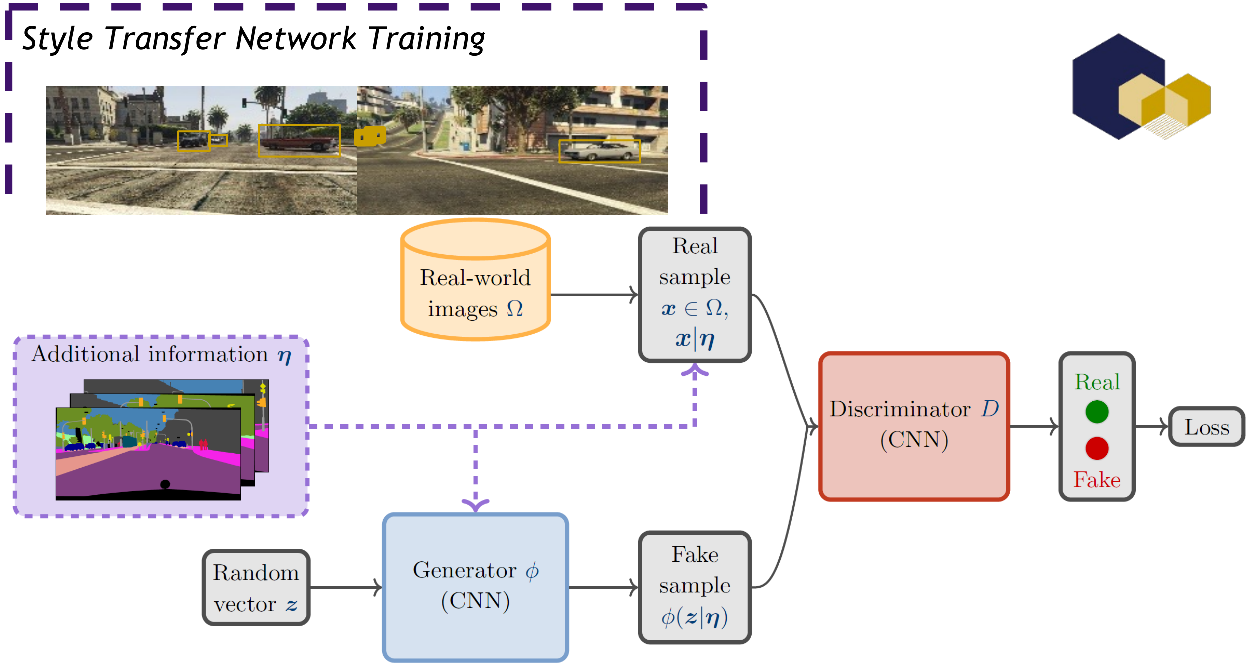
- Can we use synthetic data to **identify AI model performance issues**?
- Can we use synthetic data to **fill identified gaps**?

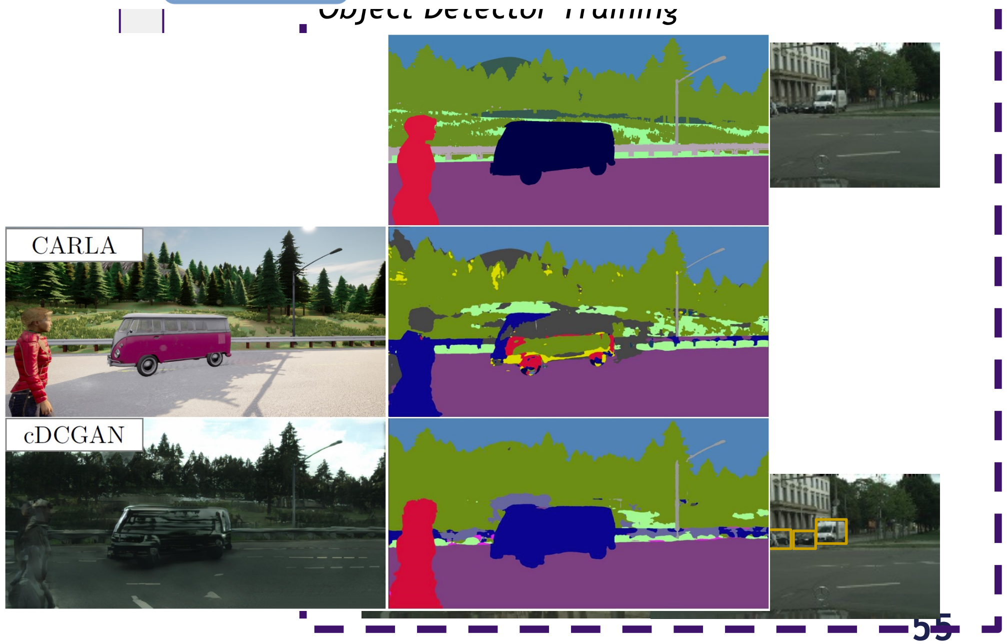


[1] Zhang, R., et al, "DDE process: A requirements engineering approach for machine learning in automated driving," in 2021 IEEE 29th International Requirements Engineering Conference (RE), 2021, pp. 269-279.

Mixed Training
Going the last synthetic mile

- Simulation only goes so far in terms of realism
- Last synthetic mile to realism?
  - Multiple techniques developed
    - Attention-Weighted Adversarial Domain Adaptation for Object Detection (AWADA)
    - conditional Deep Convolutional Generative Adversarial Network (cDCGAN)





KI DATA TOOLING

The Data Kit for Automotive AI

Mixed Training Challenge

Mixed Training
Poster and Deep-Dive Teaser

Interested? – join our **Deep-Dive**

- Training **strategies**
- Image **stylization**
- Unsupervised **domain adaptation** for object detection
- **Identifying** and **filling gaps** with synthetic data
- A novel mixed training **challenge**

Deep Dive #4
Training with Synthetic Data - Mixed Training

or visit our **posters**:

Poster Mixed Training - Identification and Filling of Data Gaps

Poster Unsupervised Domain Adaptation for Object Detection using Adversarial Style Transfer and Semi-Supervised Learning

Mixed Training Challenge

Vision: Synergetic combination (i.e., mixing) of synthetic and real data for targeted filling of gaps

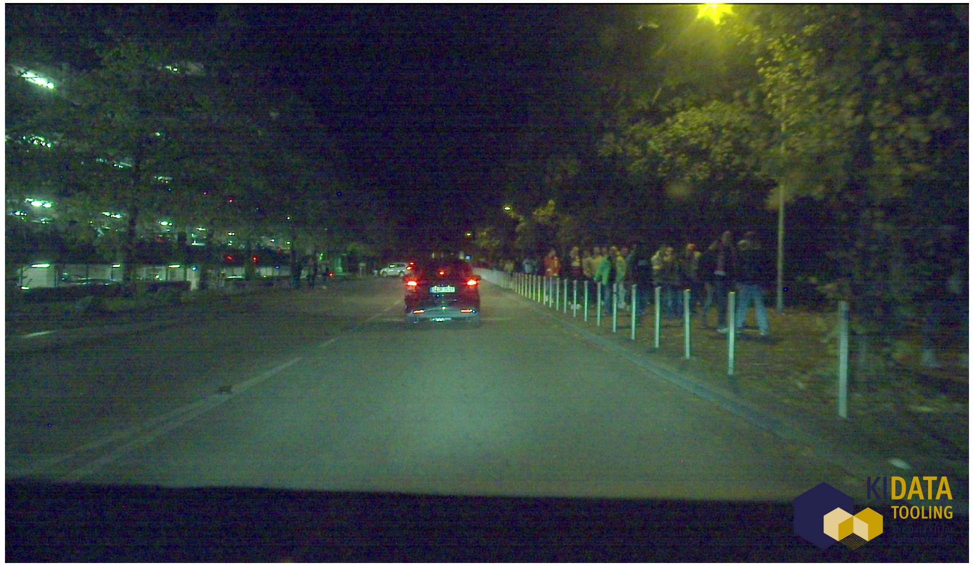
Task: Improve pedestrian detection in general and at night in particular.

Challenge:

Train a pedestrian detector in a mixed **fashion** to improve pedestrian detection at night images.

Datasets:

- Real Dataset: Data from Bosch recording vehicle
- Synthetic Dataset: Data from the KI-Absicherung project



Can you spot the pedestrians?

KI DATA TOOLING
The Data Kit for Automotive AI

Thomas Stone, thomas.stone@bmw.de

Claudia Drygala, TU Berlin | Evren Ermis, Continental | Jakob Kirner, BIT-TS |

Tobias Knerr, Uni Passau | Maximilian Menke, Robert Bosch GmbH |

Philipp Rigoll, FZI

KI Data Tooling is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

KI FAMILIE

Supported by:

Federal Ministry for Economic Affairs and Climate Action

on the basis of a decision by the German Bundestag

www.ki-datatooling.de @KI_Familie KI Familie

Deep Dives

11:20 - Deep Dive #1 | ROOM 1
Training with Synthetic Data - Mixed Training

13:00 - Deep Dive #4 | ROOM 1
Corner Case

11:20 - Deep Dive #2 | ROOM 2
Real Data Overview

13:00 - Deep Dive #3 | ROOM 2
Deep Dive synth. data production & validation