Labeling specifications

key learnings and industry trends from 4 years with OEMs
**Why**

We believe that AI is the most powerful tool our generation has at hand. We make it accessible to bring society to the next level.

**How**

We enable you to get the right data at the right quality and the right quantity to bring AI into real-world products like self-driving cars. For the benefit of everyone.

**What**

We provide a world class SaaS tool chain to select the right data, generate high quality 2D & 3D data sets and real world test scenarios at scale.

**TL; DR:** We solve the data problem of AI!
A young fast-growing start-up

understand.ai | a dSPACE company

Founded: March 10th, 2017
Research dates back to 2014

HQ: Karlsruhe (Germany)
Heart of a key European Automotive & AI cluster

Team:
53 associates, combination of AI & Automotive talent

Customers:
Automotive OEMs and Tier 1/2s
e.g. VDA V&V-Methoden (Pegasus successor)

Offerings:
SaaS Annotation tooling, Annotation Service and Scenario Extraction
Portfolio - Focus on Automation and Quality

2D Bounding Box

Semantic Segmentation

Polyline Annotation

LiDAR / Radar

State of the art SaaS tooling
Generation of simulation scenarios

**Input**
1. Camera data, LiDAR, ...
2. Object list from vehicle bus, GNSS, HD Map, ...

**Output**
1. Object trajectories for replay
2. Logic scenarios for simulation, sensor-realistic 3D scenes

**SERVICE**
- From recorded data to logic scenarios
- Based on camera data or object lists
- Applicable to other sensor inputs, e.g. LiDAR
- Independent from HD-Maps*

**Cloud**
- Pay-Per-Scenario

**Output**
- Road Network
- Maneuver
- 3D Scenery
- Traffic
Daily challenges

Working with labeling specifications
Process of an annotation projects

Expectation

- Data Acquisition
- Data Annotation
- Model Training
- Application
- Validation
Process of an annotation projects

Expectation in an ideal world

- Data Acquisition
- Data Annotation
- Model Training
- Application
- Validation

Data Annotation
Define Specifications
Get high quality labeled data
Truck or Car?

Every day challenges
Separate cars or part of the truck?

Every day challenges
Lane markings - expectation
Every day challenges
Lane markings - reality

Every day challenges
Unique traffic participants

Every day challenges
Key learnings

Goal: Consistent and high quality data sets
Best practices for defining specifications

Goal: Consistent and high quality data sets
Best practices for defining specifications

Goal: Consistent and high quality data sets
Be precise with wording

Guidelines

Class: Truck
This includes all trucks.

Class: Truck
This includes all trucks. Please include any cargo on the truck. Pick-up trucks are considered trucks as well.
Everyone understands pictures

Guidelines

Class: Truck
This includes all trucks.

Class: Truck
This includes all trucks. Please include any cargo on the truck. Pick-up trucks are considered trucks as well.
Use exceptions (edge cases) wisely

Meta attributes can split complex tasks

**Class: Truck**

This includes all trucks.

Mark the trailer of the truck separately, if you see the whole trailer. If there is a second trailer only mark the last trailer.

**Class: Truck**

This includes all trucks. ..... Set a meta attribute “occluded trailer” if the trailer is occluded by other traffic participants or objects.
Feedback pipeline

Creating a feedback pipeline is crucial for solving specification issues early on.
5 by 5 rule

Hands on solution for spotting labeling spec inconsistencies
Industry trends
From academia to industry

Shift in work focus

Source: Andrej Karpathy
Academia

Academia versus Industry

- **Training Data**: Fixed dataset for comparability
- **Model**: Variable, high sophisticated algorithms
- **Performance**: Variable results, often hard to reproduce
Industry

Academia versus Industry

Training Data

Model

Performance

**variable**, ever changing data set

**variable** but less relevant algorithms

**fixed** performance results needs to be reliably reproducible under a very wide range of different circumstances
Fixed Model Performance

Industry
Annotation projects

Industry: Past

Outcome: Non-compatible datasets in the same company
Annotation projects

Industry: Present

Outcome: compatible datasets inside a company?
Annotation projects

Industry: Future

(Some) standardised Datasets:

+ Lower entry barrier
+ Best practices integrated
+ High quality
+ High quantity
+ Reasonable cost

Outcome: compatible datasets between companies
Quality annotation meets scenario

UAI | dSPACE Scenario Generation Service