Empowering the Application of Machine Learning Techniques through Simulation Data Management

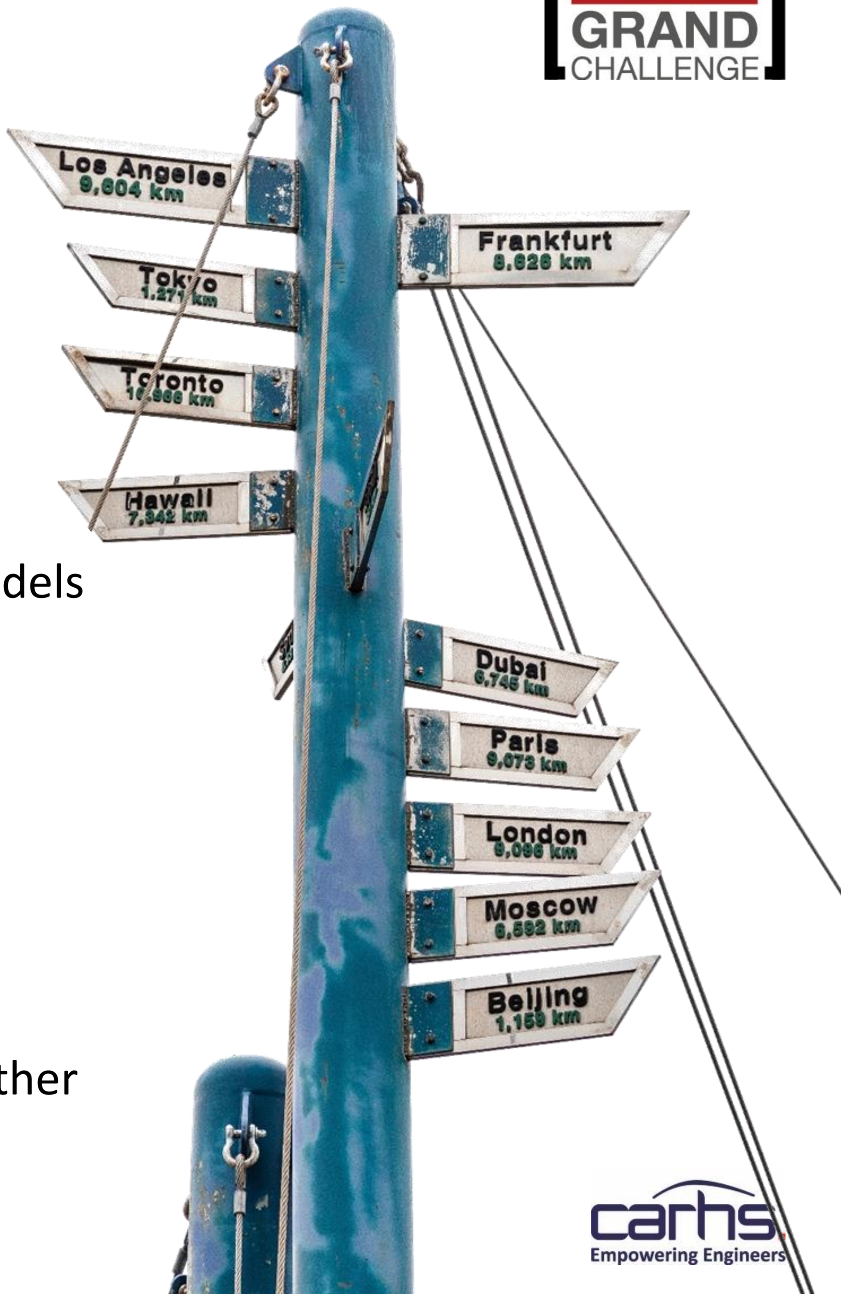
Marko Thiele – SCALE GmbH;
Akhil Pillai – TU Dresden; Prof. Dr.-Ing. habil. Uwe Reuter – TU Dresden

automotive CAE Grand Challenge 2019
April 16 - 17, 2019
Hanau, GERMANY

automotive CAE GRAND CHALLENGE

Agenda

- *Challenge #1: Provide simple and fast data access*
  Enabling simple and fast access to large amounts of Data
- *Challenge #2: Handle arbitrary simulation data with acceptable performance and robustness*
  Handling of many and also sparse data with acceptable performance
- *Challenge #4: Ensure process chain reliability, stability and fault tolerance*
  Implementing fault tolerant processes
- *Challenge #5: Educate engineers & students to get skills beyond scripting*
  Empowering engineers to develop simulation processes and not just models
- *Challenge #3: Provide reliable applications understandable by non data-scientists*
  Usability of applications that empower machine learning techniques
  - Usability of SDM systems
  - Example: automatic spot weld generation
- *Challenge #6: Provide benchmark data to enable algorithm development and comparison*
  Development of algorithms with limited access to data
- *Challenge #7:*
  Creating an environment where all the data is actually related to each other

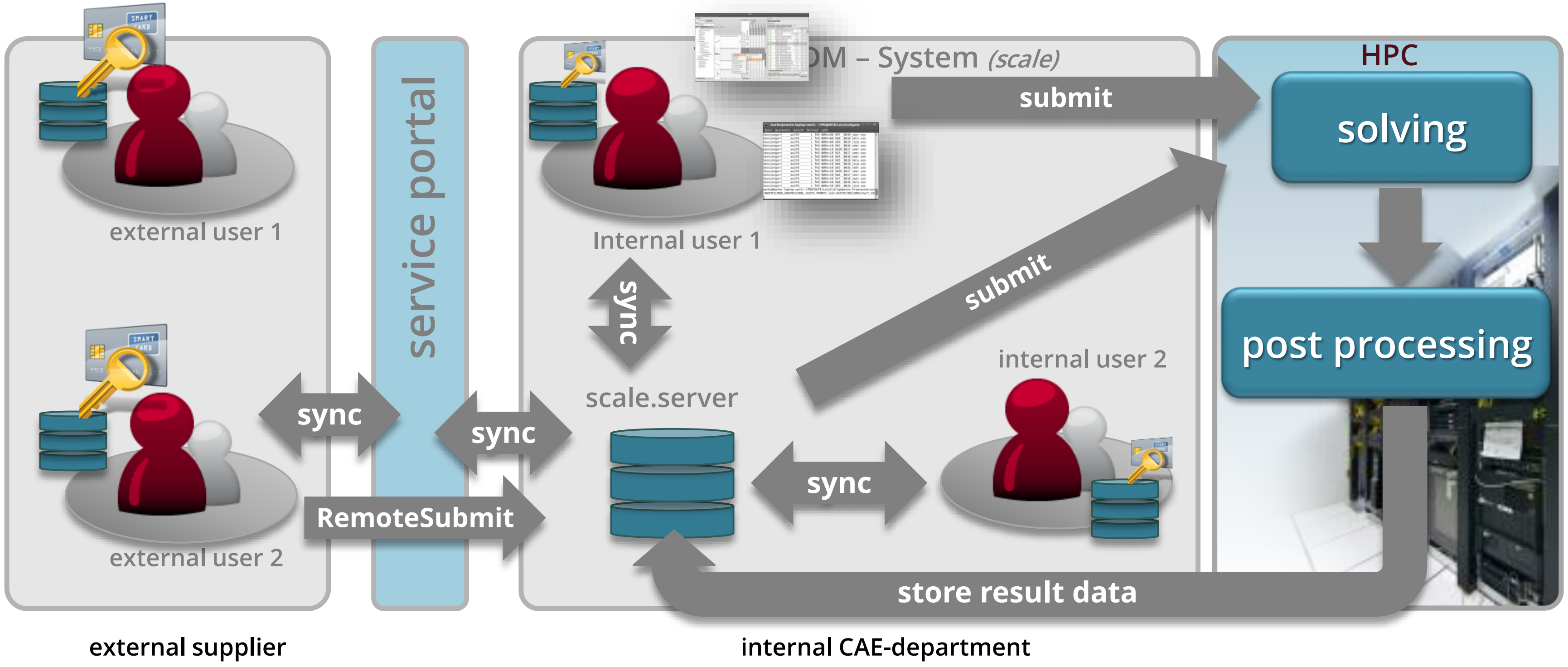


carhs
Empowering Engineers

Challenge #1: Provide simple and fast data access

Enabling simple and fast access to large amounts of Data

- Requirements
  - Uncomplicated
  - Fast
- Challenges
  - Security
  - Data volume
  - Amount of data
  - Heterogeneous work environment
  - Collaboration



Sync
decentralized

- Central data storage, synchronization with local workstations *(cloud like infrastructure)*
- Encrypted transfer, encrypted storage *(two factor authentication and encryption)*
- Offline handling of data *(RichClient)*

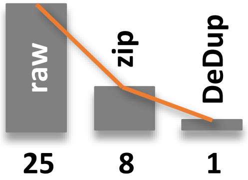
Offline / Online
performance

- Users/Teams are independent of servers and infrastructure
- Users work with local data
- Good performance while application of preprocessing tools

Challenge #2: Handle arbitrary simulation data with acceptable performance and robustness

Handling of many and sparse data with acceptable performance

- Requirements
  - Open files fast
  - Work with many files
  - Solver independence
- Reduction of data for transfer and storage
  - **Inputs:** Deduplication

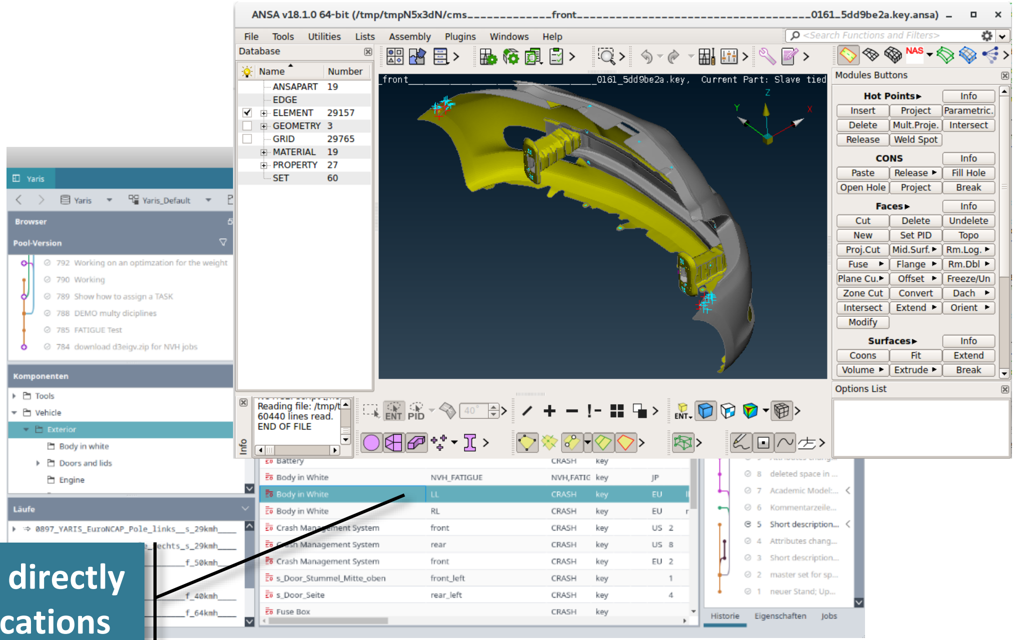


} factor of ~8

- **Results:** SDMZIP factor of ~3

SIDACT GmbH

- Challenges
  - FEA files are sparse
  - It's not known which files the user opens next
  - Many simulation results
  - Many different tools
- Desktop integration
  - Instant access to data
  - Fast loading
  - Tight integration of FEA tools



open data directly into applications

SCALE

carhs. Empowering Engineers

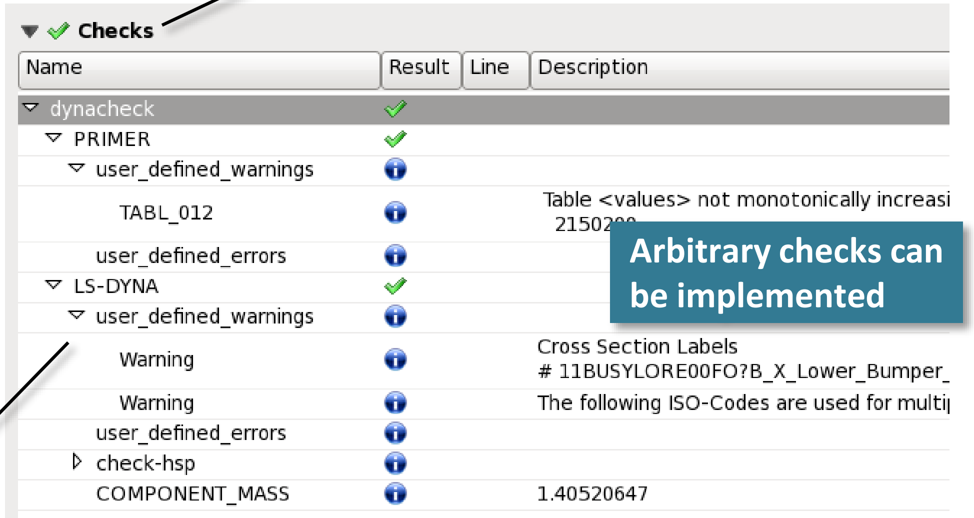
Challenge #4: Ensure process chain reliability, stability and fault tolerance

Implementing fault tolerant processes

automotive CAE GRAND CHALLENGE

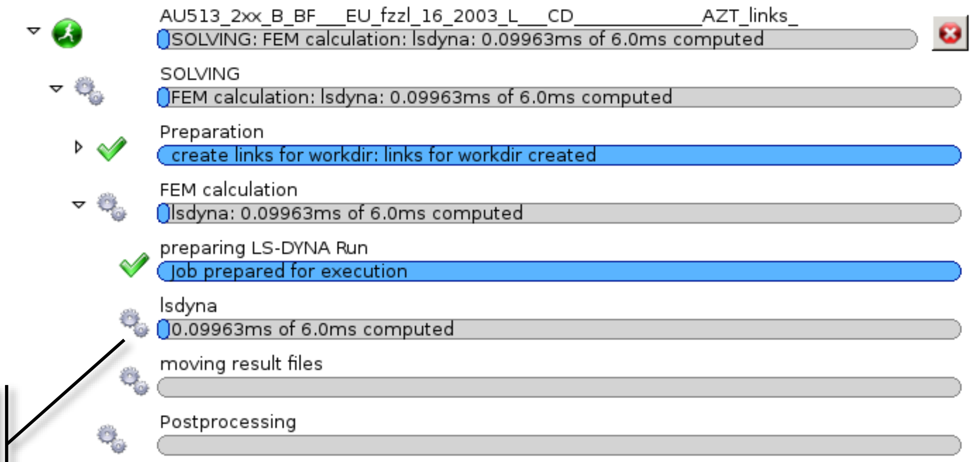
- Requirements
  - Each simulations should return valuable results
  - Model errors should be detected upfront
  - Process errors should be easy to understand and fix
- Challenges
  - Complexity of simulation models
  - Sophisticated FEA methods
  - Complex HPC environments
  - Number of people working on one model

Every time something is saved it will be checked



Arbitrary checks can be implemented

Check results pop up immediately



Tight monitoring of each process step

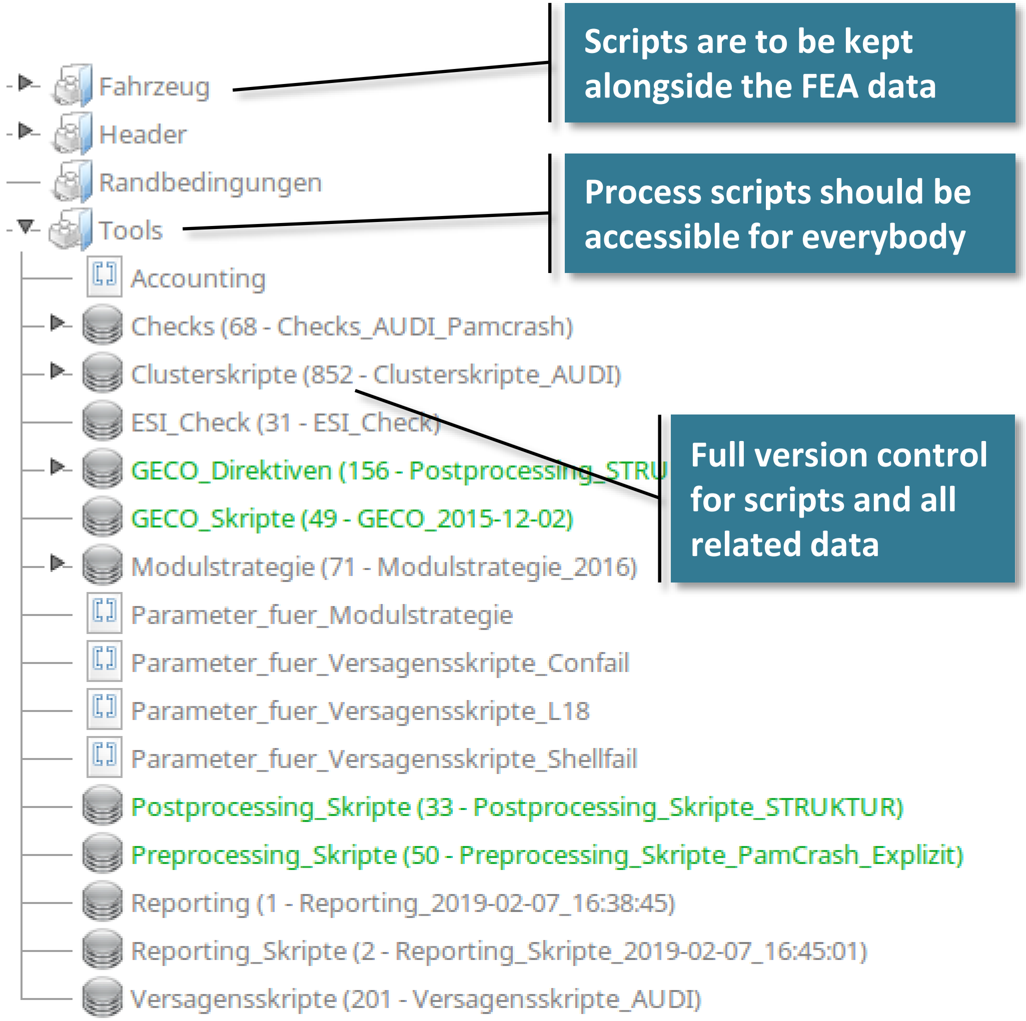
- Continuous testing of model data
  - Each time a change is applied
  - Before each simulation
  - Customizable
- Tight process monitoring
  - Instant feedback to users
  - Easy to understand error messages
- Automated testing for processes
  - Unit testing for process scripts

SCALE carhs Empowering Engineers

Challenge #5: Educate engineers & students to get skills beyond scripting

Empowering engineers to develop processes and not just models

- Requirements
  - Programming should be made easy
  - Integration into normal workflows
- Challenges
  - Education
  - Multitude of programming tools
  - Teamwork



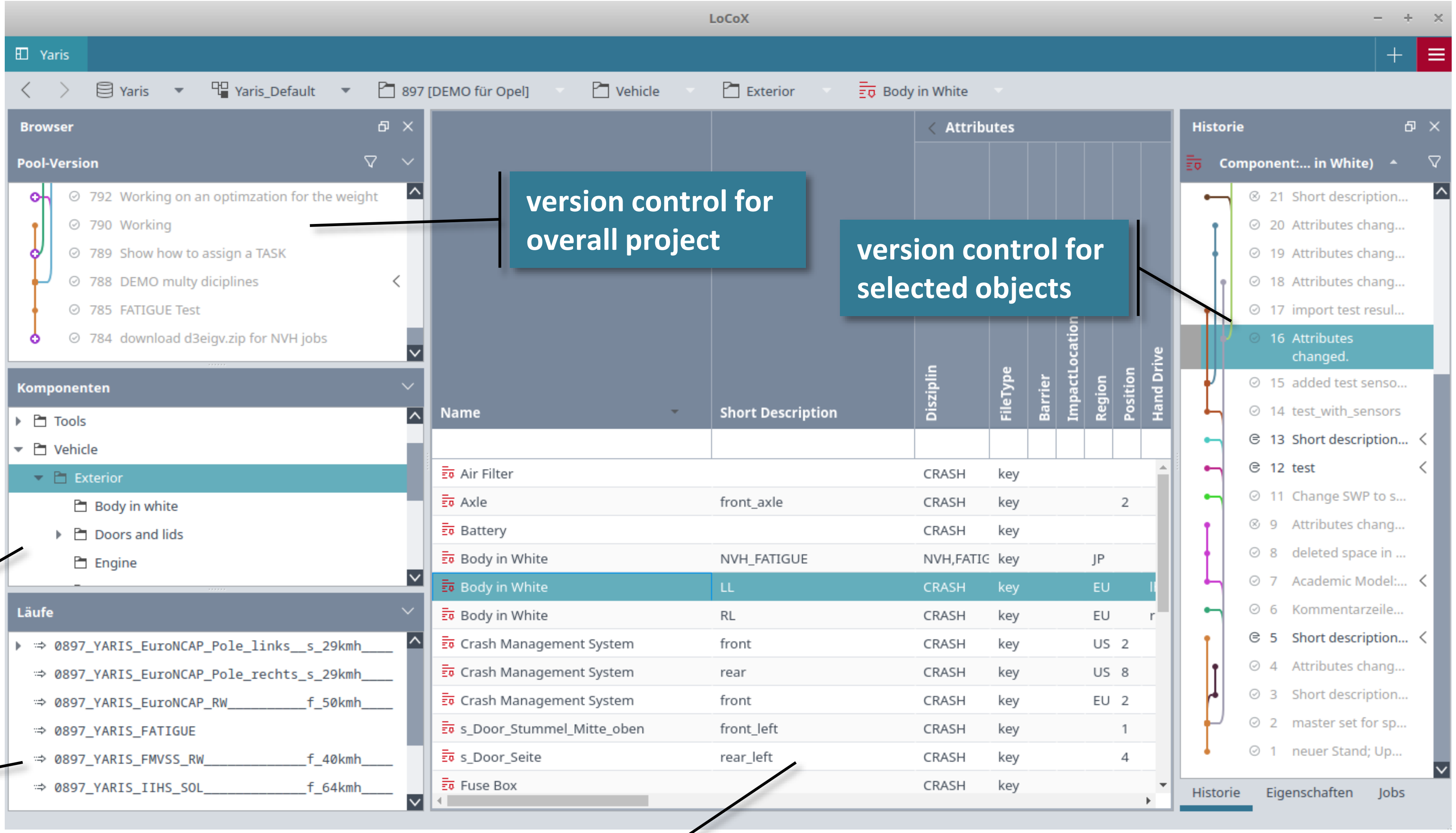
- All data should be exposed by APIs

Challenge #3: Provide reliable applications understandable by non data-scientists

Usability of applications that empower machine learning techniques

automotive CAE GRAND CHALLENGE

- Requirements
  - Intuitive User interface
  - Responsiveness
- Challenges
  - Complexity vs. intuitiveness
  - Huge data amounts



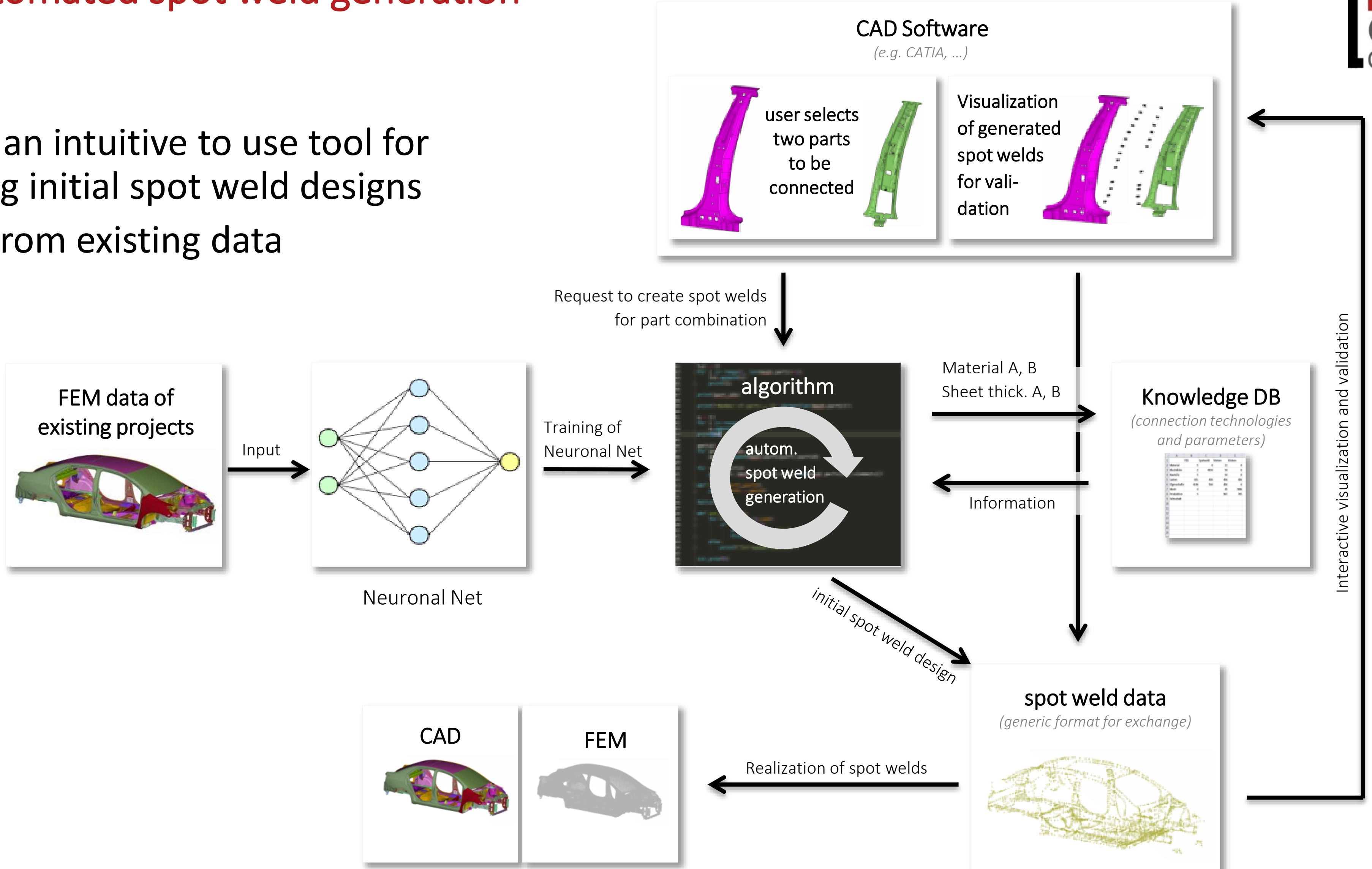
SCALE

carhs. Empowering Engineers

Challenge #3: Provide reliable applications understandable by non data-scientists

Example: automated spot weld generation

- Goal
  - Create an intuitive to use tool for creating initial spot weld designs
  - Learn from existing data



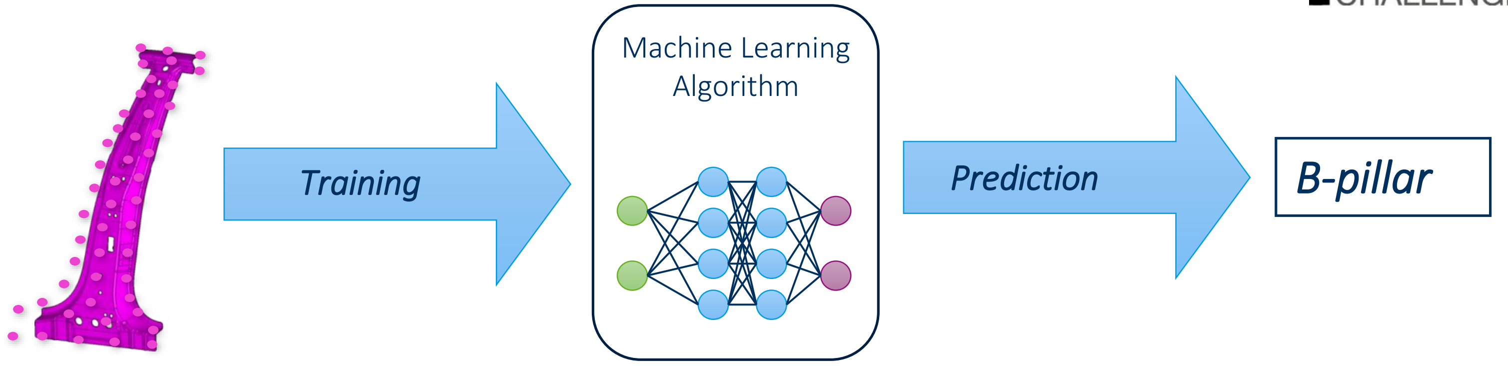
Challenge #3: Provide reliable applications understandable by non data-scientists

Example: automated spot weld generation

1st part recognition

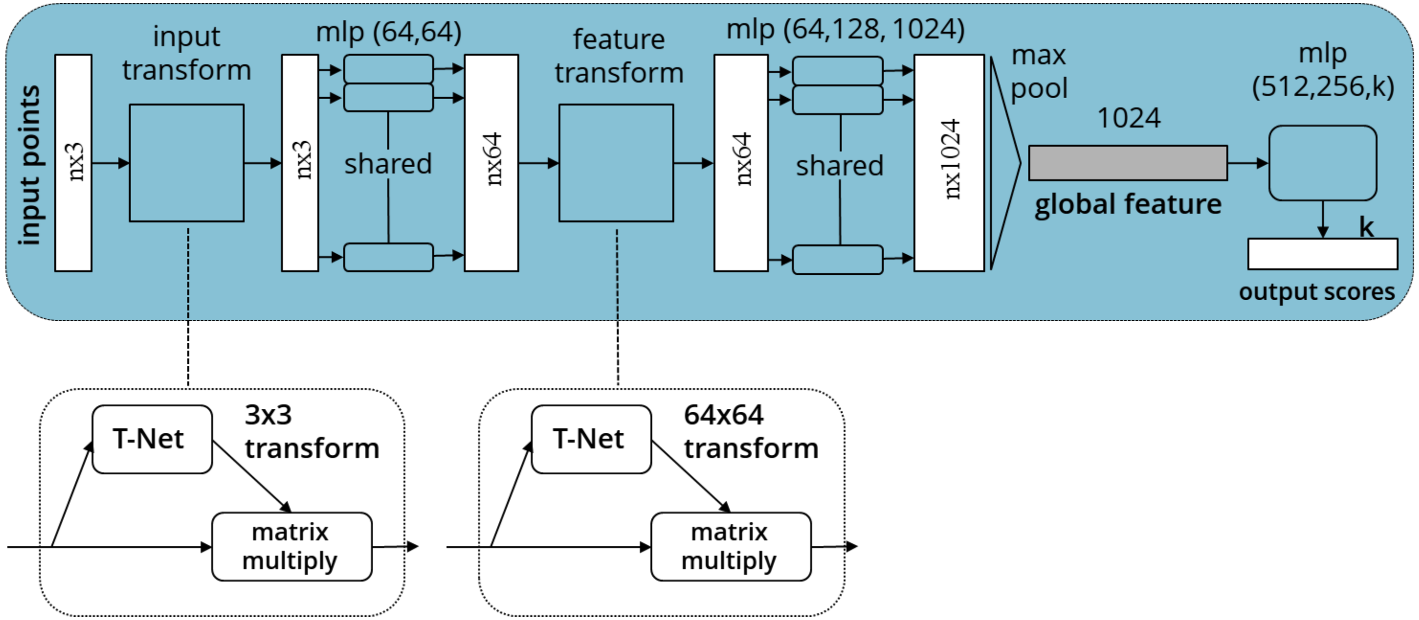
(classification problem)

- Independent of geometric representation
  (CAD, FEM, …)
- Robust to
  - slight geometric differences
  - orientation
  - spatial location
- State of the art recognition accuracy



Point cloud

PointNet[1]



[1] PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation
Charles R. Qi, Hao Su, Kaichun Mo, Leonidas J. Guibas

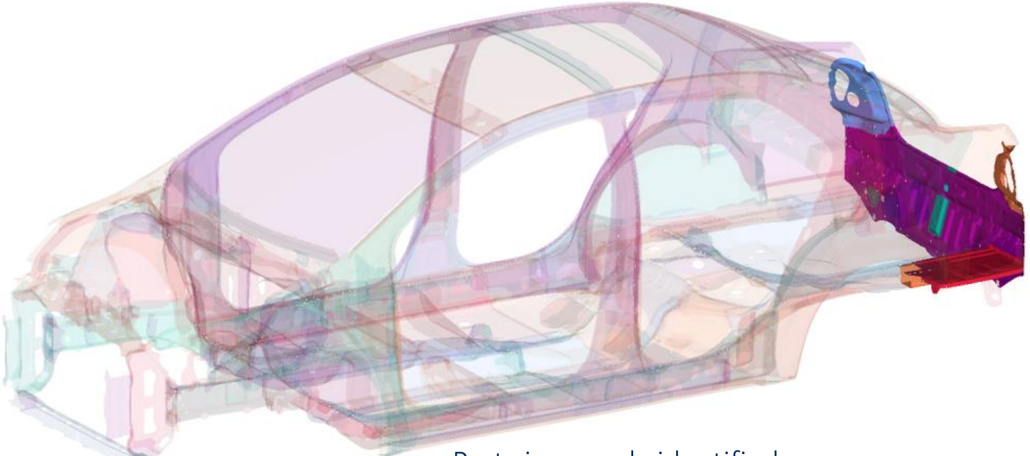
TECHNISCHE UNIVERSITÄT DRESDEN SCALE carhs Empowering Engineers

Challenge #3: Provide reliable applications understandable by non data-scientists

Example: automated spot weld generation

Part Identification of TOYOTA YARIS Model

Number of parts	250
Number of points per part	1024
Training samples	10,000
Test samples	2500
Epochs	300
Training time	~30 hours

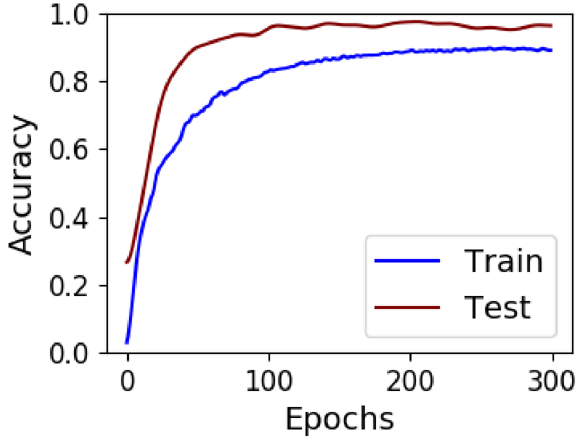


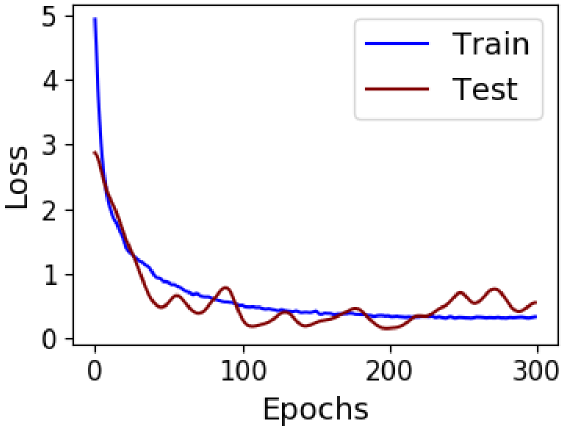
Parts incorrecly identified

- 237 parts identified correctly
- 95 % identification accuracy

✓ Part identification

⚠ one version of part only



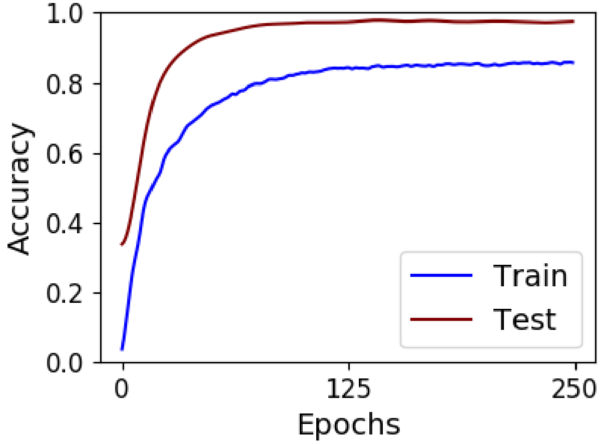


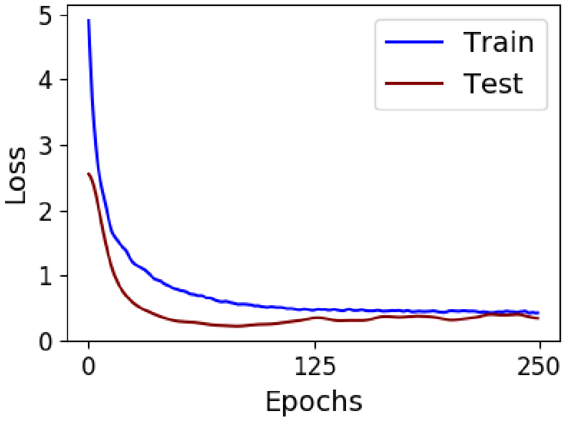
Challenge #3: Provide reliable applications understandable by non data-scientists

Example: automated spot weld generation

Part Identification of AUDI Model

Number of parts	350
Number of points per part	1024
Training samples	17,500
Test samples	5000
Epochs	300
Training time	~48 hours





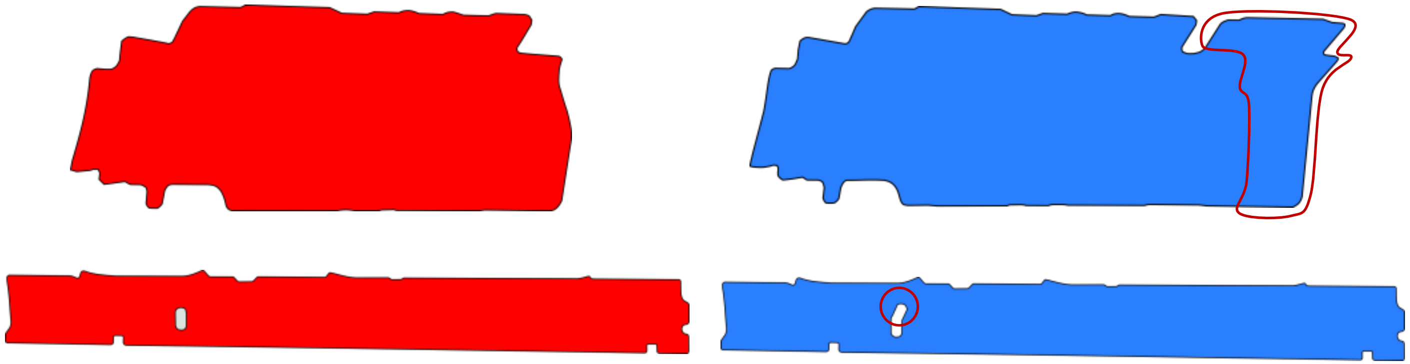
- 339 parts identified correctly
- 96 % identification accuracy
- New version of 13 parts used for evaluation

✓ Part version identification

⚠ Parts of one model only

Examples of part versions

Training geometry	New version



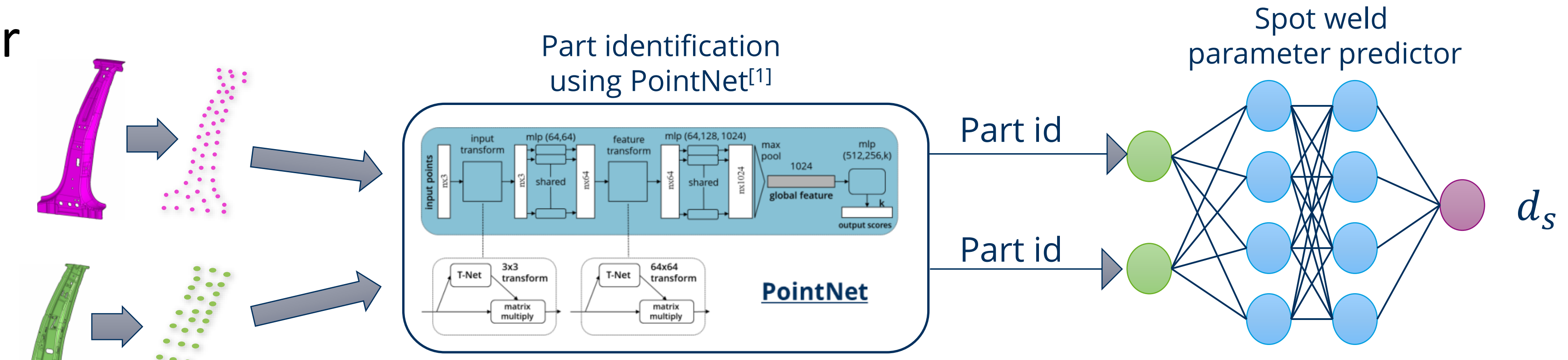
Challenge #3: Provide reliable applications understandable by non data-scientists

Example: automated spot weld generation

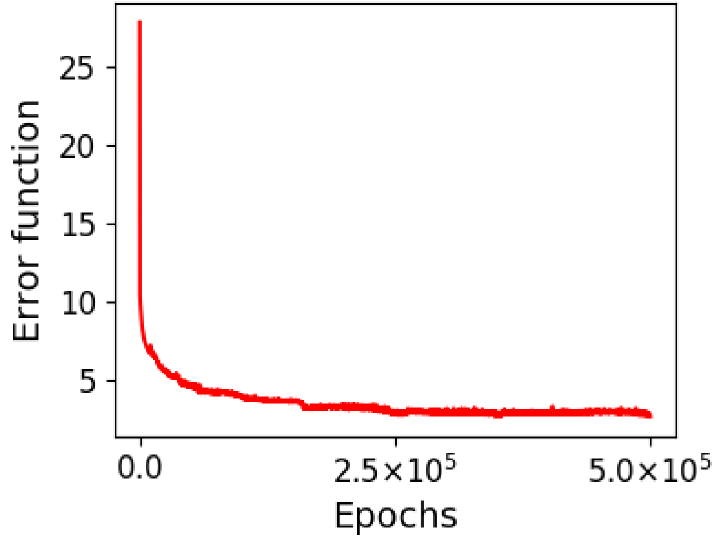
Estimation of spot weld design parameter

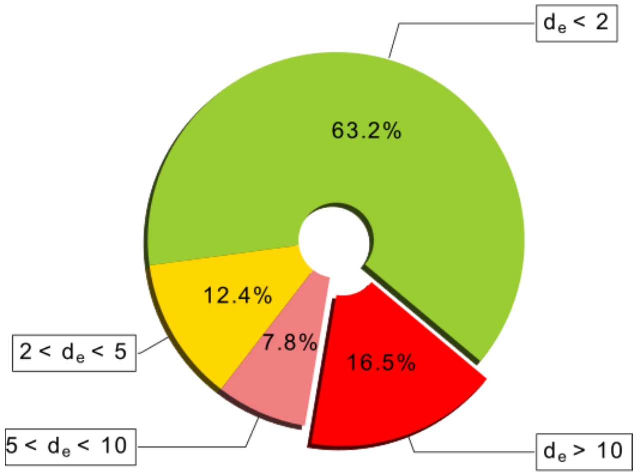
(prediction based approach)

- Support for geometric and non geometric information
- Extendable to more complex spot weld design parameters



- Architecture : 2 | 150 | 50 | 1
- Error function: Mean Squared error
- Gradient descent optimization



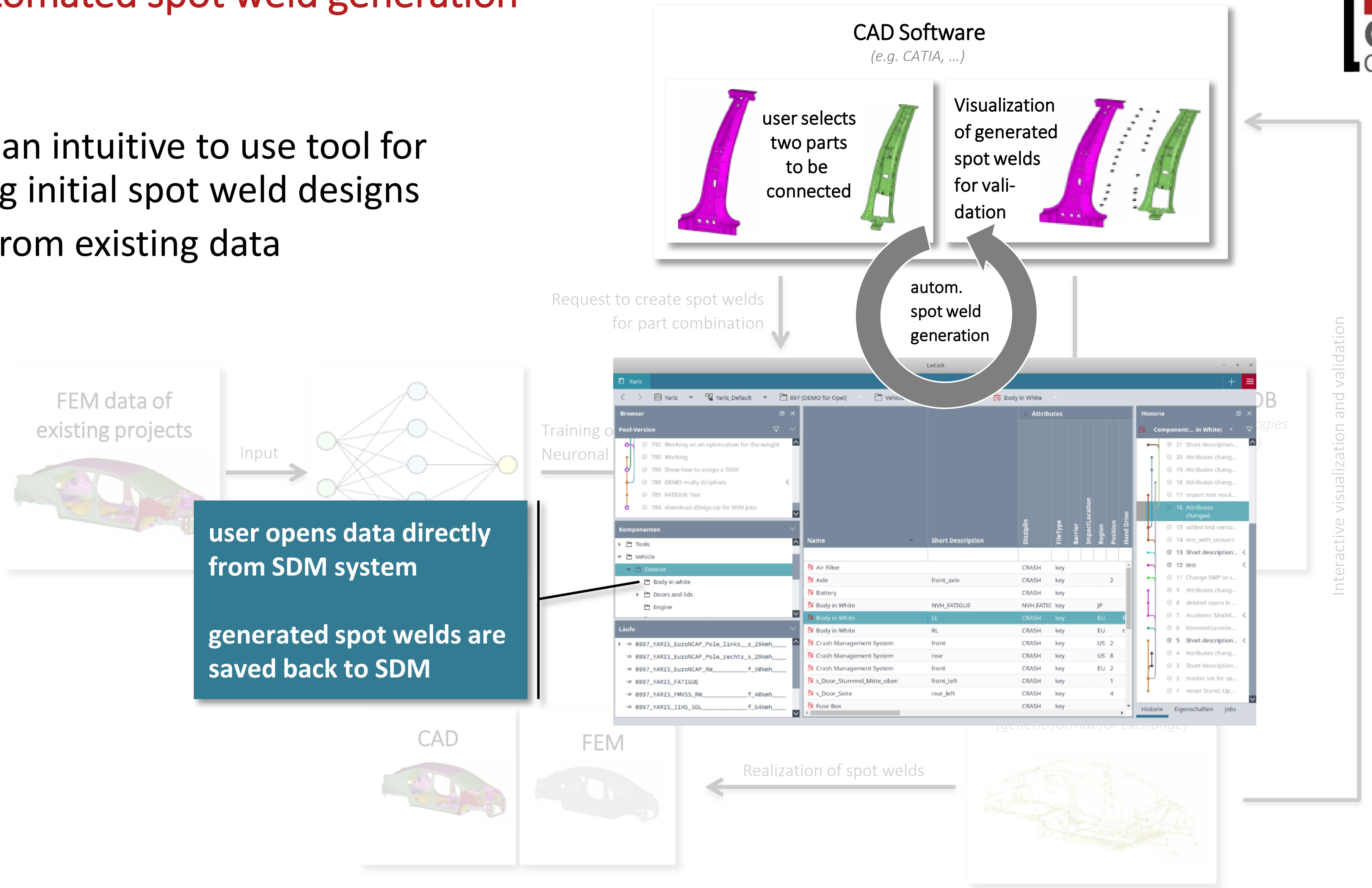


[1] PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation
Charles R. Qi, Hao Su, Kaichun Mo, Leonidas J. Guibas

Challenge #3: Provide reliable applications understandable by non data-scientists

Example: automated spot weld generation

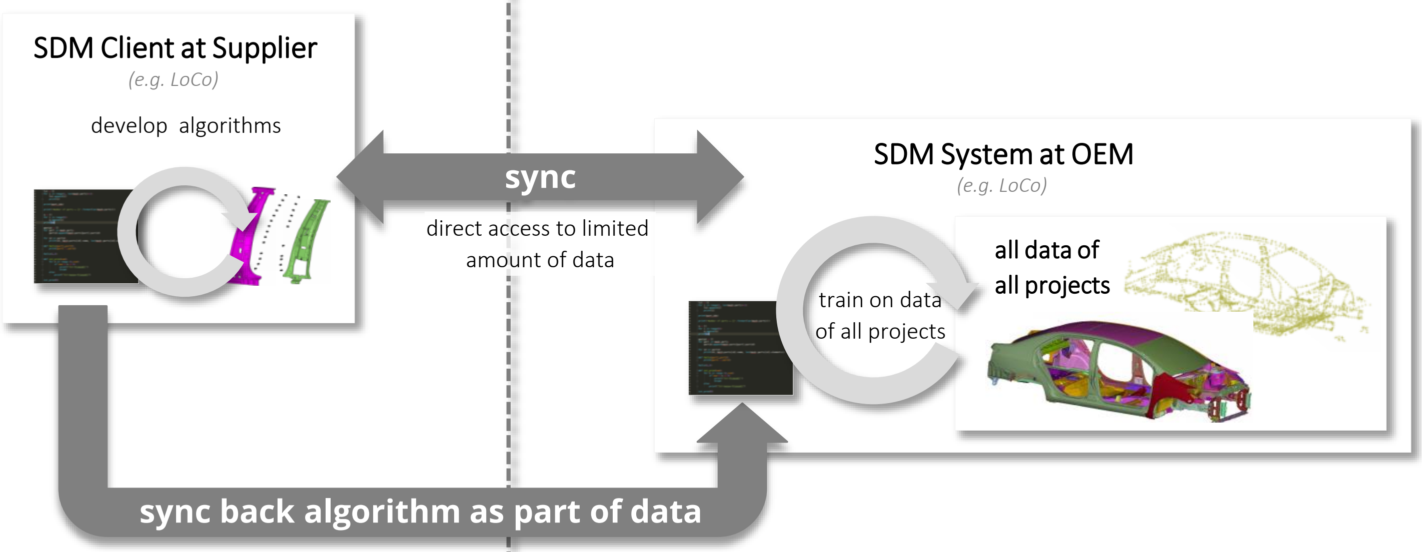
- Goal
  - Create an intuitive to use tool for creating initial spot weld designs
  - Learn from existing data



Challenge #6: Provide benchmark data to enable algorithm development and comparison

Development of algorithms with limited access to data

- Requirements
  - Data is needed to develop algorithms
  - Data sets for training should contain as many data as available
- Challenges
  - OEMs want to protect their IP
  - Amount of data that would have to be transferred
  - Resources for training of data
  - Security requirements



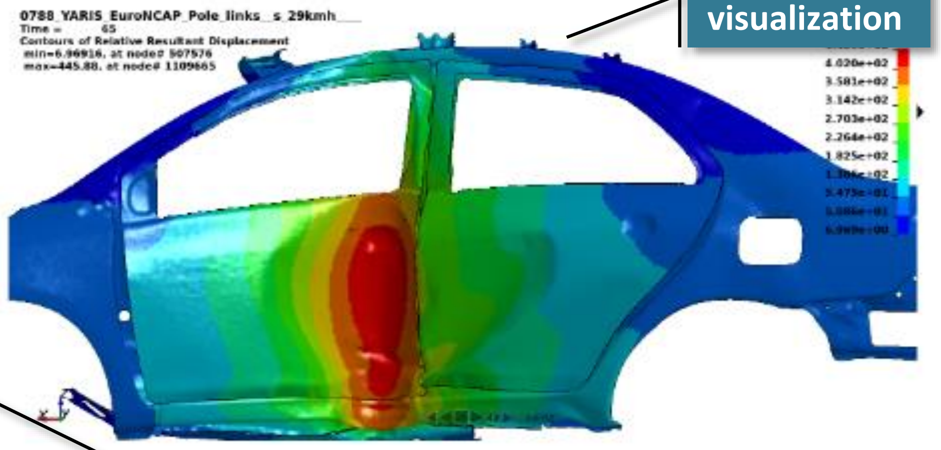
SCALE carhs. Empowering Engineers

Challenge #7:

Create an environment where all the information is related to each other

- Requirements
  - All knowledge needed to develop machine learning algorithms has to be captured
  - Relations between applied changes and observed effects are to be captured
  - Assessments for applied changes
- Challenges
  - Implementation at OEMs is challenging
  - Human users are needed for creating assessments
  - “Chicken or egg” problem…
    there is currently no such related and/or assessed data

interactive visualization



autom. detection of changed behavior

SDMZ SIDACT GmbH

Typ: automatic detection
Tags: displacement
Evaluation: 9,5%
Event Name: ...
Event ID: 123

Description

increased intrusions

tags for measures related to privous similar events

Tag	Frequency	Rating (better, worse)	
increase thickness	62%	63%	27%
remove part	42%	58%	42%
material	35%	43%	57%

suggestions based on previous assessments

SCALE carhs. Empowering Engineers

Conclusions

- **Simulation Data Management** is the basis for gathering and providing the data needed for Machine Learning Applications

- Creating **truly related data** in the right format and with the correct amount of **data reduction** is essential

- Application of **machine learning** approaches will help to **solve many "small" problems** of the whole picture

- Current machine learning approaches can already be **very valuable** and **many things remain to be done** in order to harvests the full potential offered by the field of artificial intelligence. However we are still far away from strong artificial intelligence*.

  *(*there won't be any time soon an automat that's able to take over the entire development of cars from us and be better in it than we humans)*

so long, and thanks for all the fish ...