Supported by:

Federal Ministry
for Economic Affairs
and Climate Action

on the basis of a decision
by the German Bundestag

KIDELTA LEARNING
Scalable AI for Automated Driving

KI FAMILIE

KI Delta Learning 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.

www.ki-deltalearning.de

@KI_Familie

KI Familie

KI Delta Learning

Autonomy at Scale

Scalable AI for Automated Driving

KIDELTA LEARNING
Scalable AI for Automated Driving

Welcome

Dear Readers,

As coordinator of KI Delta Learning, I am delighted to present the results of this project. The booklet you are holding in hands shows the achievements of more than three years of intensive work in a nutshell. More than 300 people teamed in 6 sub-projects and 21 work packages to advance autonomy at scale. This team effort has paid off!
On the following pages, we are introducing our project topics, illustrating our approaches and presenting around 50 research topics.
17 project partners and four external partners collaborated, shared findings and extended their knowledge mutually. Our partners include OEMs, automotive suppliers, technology providers as well as universities and research institutes. This mix of partners allowed a quick transfer of results and approaches from different research and technology fields to the industry. In return, academia received a better understanding of requirements and challenges regarding product development.
During the last three years, our project published more than 90 scientific papers and established the workshop Autonomy@Scale at IEEE Intelligent Vehicles Symposium.
Initiated by VDA Leitinitiative Connected and Autonomous Driving, KI Delta Learning would not have been possible without the commitment of all partners involved.
Their different expertise and backgrounds were the basis of our common success.
The support and guidance of the Federal Ministry for Economic Affairs and Climate Action as well as the project officer TÜV Rheinland Consulting were helpful mastering the various project phases. On behalf of the whole team, I would like to thank them very much.
I would like to thank the deputy lead ZF and the other sub-project leads from Valeo, Bosch, University of Wuppertal and DLR as well as the project management team at EICT for their valuable contributions in coordinating the project.
My special gratitude goes to all of you, who contributed to the great results of KI Delta Learning. All of you, who have been active to set up, manage, fund and research. You made this project a true success and a pleasure to work in. Autonomy at scale took an important step on the way to reality!

Dr.-Ing. Amin Hosseini
Project Coordinator
Mercedes-Benz AG

Greeting from the Federal Ministry for Economic Affairs and Climate Action

Autonomous and connected driving will shape the mobility of the future and enable completely new concepts, including aiming towards the goal of climate neutrality. There are numerous research and development issues to be addressed along the way. Traffic is not only diverse and complex, but also constantly changing. Highly and fully automated vehicles must be able to handle a wide variety of situations safely and reliably. This results in one of the most challenging and exciting application areas for artificial intelligence. With its flagship projects from the “KI Familie”, the Federal Ministry for Economics Affairs and Climate Action is promoting a cooperative research approach that brings together the expertise of numerous participants. This yields an outstanding foundation for the next step toward the safe implementation of broad AI know-how in vehicles. Germany must keep addressing the central questions surrounding the future of autonomous mobility in order to assert itself in international automotive competition.

Over the 39 months duration of the KI Delta Learning project, methods and tools were developed that enable a more efficient training of AI and render the unrestricted use of automated vehicles in the „Open World“ possible. The methods of versatile machine learning that were explored, are setting a new standard for a more efficient training of AI. This marks the end of the second of four projects of the “KI Familie” that were jointly launched in 2019. We would like to thank all participating partners from industry and science for their outstanding work and results. They have thus provided another important building block on the way to the unrestricted use of automated vehicles!

Ernst Stöckl-Pukall
Head of Division for Digitalisation and Industry 4.0, Federal Ministry of Economic Affairs and Climate Action

Federal Ministry
for Economic Affairs
and Climate Action

Collaboration in Artificial Intelligence

Classic automotive questions are re-emerging with regard to AI. AI technology know-how and its safe use in modern vehicles will determine the leading role in the mobility markets of the future. The German automotive industry addresses this challenge with the projects of the **KI Familie**. The KI Familie was initiated and developed by the **VDA Leitinitiative Connected and Autonomous Driving**. 80 leading partners from science and industry are involved receiving funding from the Federal Ministry for Economic Affairs and Climate Action (BMWK).

In this unique setting, all KI Familie projects are working together. The partners are sharing knowledge while fostering pre-competitive collaboration which is essential in an ever more competitive and complex environment with fast pace innovations. Exchanging findings across project boundaries accelerates the knowledge buildup in cutting edge technologies for the good of industries, research institutions and society. The joint commitment to share pre-competitive knowledge helps each partner to stay technologically ahead and multiplies resources and investments of each partner.

The KI Familie has four sibling projects which all are focusing on special AI topics.

KI ABSICHERUNG
Methods and measures to safeguard AI-based perception functions for automated driving.
https://www.ki-absicherung-projekt.de

KI WISSEN
Development of methods for the integration of knowledge into machine learning.
https://www.kiwissen.de

KI FAMILIE

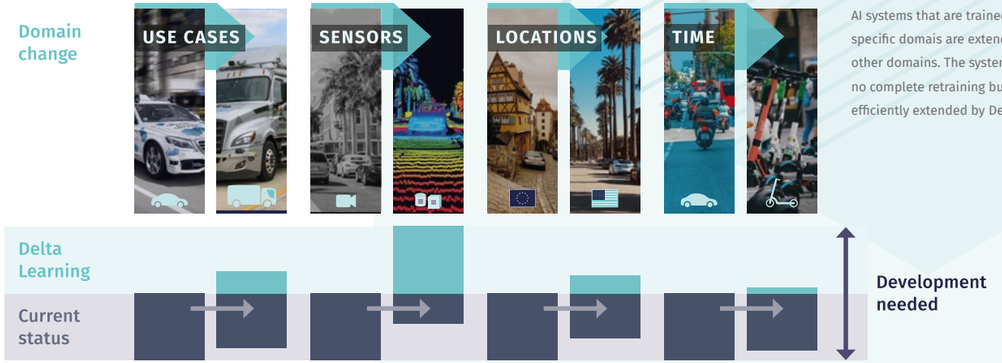
KI DELTA LEARNING
Development of methods and tools for the efficient expansion and transformation of existing AI modules in autonomous vehicles to meet the challenges of new domains.
https://www.ki-deltalearning.de

KI DATA TOOLING
Methods and tools for the generation and refinement of training, validation and safeguarding data for AI functions in autonomous vehicles.
https://www.ki-datatooling.de

KI Delta Learning

Scalable AI for Automated Driving

Highly and fully automated vehicles are facing a large variety of complex situations in a continuously evolving world of mobility. Especially for environment detection, Artificial Intelligence is a key technology. In recent years AI made huge progress, however, automotive AI was trained for limited scenarios only. To work in other environments, AI algorithms needed re-training for new domains, resulting in enormous development costs. The KI Delta Learning project has investigated new approaches in machine learning to enable more efficient training of AI modules. In turn, this leads to better adoption and more effective deployment of automated functions for the Open World.
The project aimed at bridging deltas - different requirements between a familiar domain and a new target domain. New methods to transfer existing knowledge to new application areas have been studied, developed, applied and now lay the basis for autonomy at scale.
The project efforts focused on transferring knowledge to new target domains, training methods enabling AI to learn additional requirements of changing application areas and how to adapt new techniques to automotive constraints.



AI systems that are trained on a specific domais are extended to other domains. The system needs no complete retraining but can be efficiently extended by Delta Learning.

The deltas included:

- Changes in sensors
- Divers traffic areas - from country roads to complex city traffic
- different countries
- Different daytimes, seasons and weather conditions
- Long-term traffic changes by new mobility concepts and road users
- Ongoing development of AI methods such as better training strategies and more efficient neural networks.

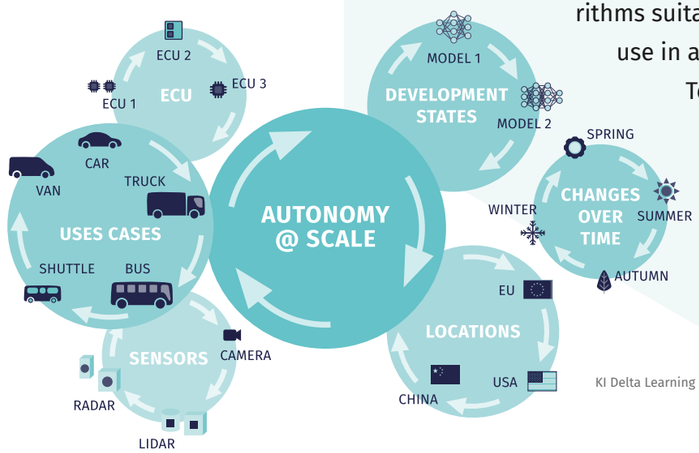
Key Facts

Starting Position and Challenges

AI modules in autonomous driving applications are scalable to a limited extent only. In fact, they are reliable only in limited domains. This results from the applied training strategies. Re-training the algorithms for each new domain, is a costly procedure. KI Delta Learning investigated disruptive methods for AI training allowing continuous learning in a more sustainable way. Knowledge that has already been learned as well as previously tested and secured development levels are retained when changing domains. This represents an efficient approach to keep up with ever shorter innovation cycles and the challenge of constantly changing mobility systems. KI Delta Learning developed methods to help closing these current gaps that limit the Technology Readiness Level (TRL) of autonomous vehicles and slow down a broad application of AI in autonomous driving.

Our Objective: Autonomy at Scale

AI used in autonomous vehicles must be responsive to a constantly evolving market and scalable to meet changing requirements. Typical examples of domain changes - deltas - are different sensors as well as changes in time and location.



KI Delta Learning Deltas, Domains and Changes.

Project Objectives – Bridging Deltas

To address these deltas, the project focused on three main areas for delta learning: transfer learning, didactics and automotive suitability. KI Delta Learning tested different approaches and aspects of these areas in order to create the next generation of AI algorithms suitable for an unrestricted use in autonomous vehicles.

To provide a basis for development in the three areas, a project-specific data set tailored to the project objectives was produced and labelled

Facts & Figures

Dr.-Ing. Amin Hosseini

Mercedes-Benz AG
Project Coordinator

40 Months

Project Duration: (01/01/2020 - 30/04/2023)

17 Project Partners

9 Industry Partners
6 Academic Partners
2 Research Institutes

€ 26,15 M

Project Budget

€ 15,87 M

Funding Budget

Funding Body

Federal Ministry for Economic Affairs and Climate Action (BMWK)

Data

Motivation

The basis for developing artificial neural networks is the underlying data. Combining the requirement to cover all deltas explored in the project and being compliant to data protection is a demanding challenge. We addressed this problem by creating a real world and a synthetic reference dataset.

Real World Recordings

First, a recording vehicle was equipped with different reference and serial sensors, co-calibrated and synchronized. Secondly, we initiated a recording campaign stretched over nearly one year. To reach a high variety, we designed different routes to cover:

- multiple countries (Germany & Italy) and cities in urban and rural areas
- different daytimes, seasons, weather and lightning conditions
- features like e-scooter and special

The quality of all recorded raw data was evaluated and a subset of frames was selected for labeling. After anonymizing all frames,

Research Vehicle at Test Field Lower Saxony

they were annotated with semantic segmentation and 3D bounding boxes. The result are 7,000,000 raw frames per sensor (193 hours). Of those, 18,000 frames have been selected to form the KI-Delta Learning dataset.

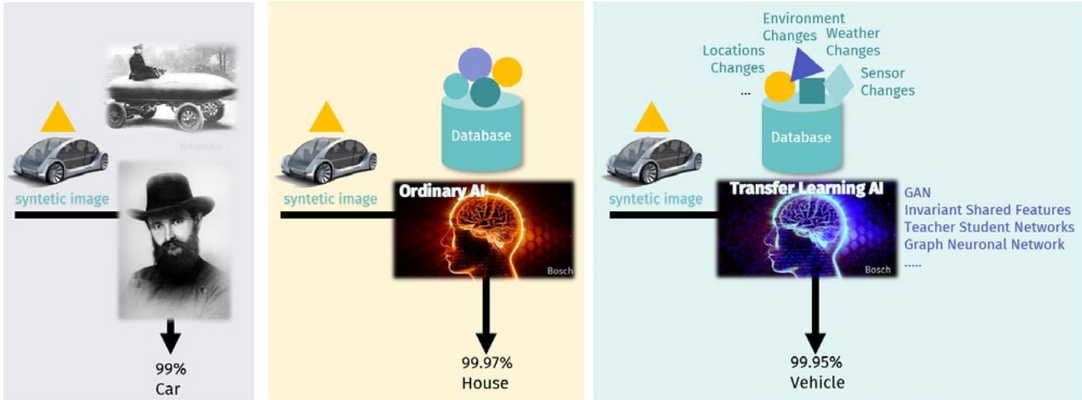
Synthetic Data Generation

In order to complete the dataset, we investigated methods to generate synthetic data. One approach was based on the CARLA Simulator, where we implemented a new serial LiDAR sensor model and improved the generation of semantic segmentation. Another approach used motion capture to generate realistic human poses and motion. To improve the quality of interaction with the virtual environment, motion capture was combined with Virtual Reality, where the actor can see not only the virtual environment, but also a body representation allowing self-perception. Reality, where the actor can see not only the virtual environment, but also a body representation allowing self-perception.

Image Generated in CARLA Simulator

Transfer Learning

In KI Delta Learning, we use transfer learning, a powerful machine learning technique that enables the reuse of knowledge acquired from one domain to another. This technique has been gaining importance in the automotive field, as it allows for the utilization of existing knowledge from other domains to improve the accuracy of autonomous vehicle systems. The partners of KI Delta Learning develop transfer learning methods, that improve the performance of vehicles with multimodal sensor equipment such as cameras, radar and LiDAR by transferring knowledge from existing data sets. Their methods can be used to reduce the amount of new data needed for training, enabling faster and more accurate predictions. Partners can quickly apply the knowledge from the existing dataset to the new problem and develop a model that is tailored to their specific needs. This can reduce the time and cost associated with developing a model from scratch and allow the partner to quickly deploy a model tailored to their specific needs. The picture on the right gives a first impression of how AI methods with transfer learning differ from ordinary AI methods. The following pages present partner examples of transfer learning methods in the automotive context.



Humans are capable of abstraction and can **abstract to a new example with the help of a few learning examples**, which may be from different domains. In this case real and **synthetic (triangle)** examples.

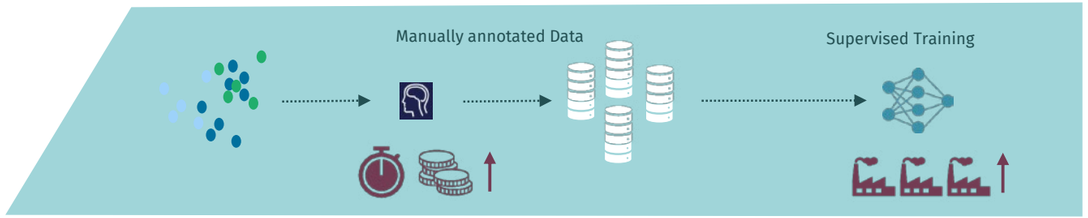
Modern AI methods work with a database. In this case, the database contains only **real examples of images (circles)**. Although a **similar image (orange)** sample is present in the data, the AI cannot abstract to the synthetic image that comes from a **different image distribution (triangle).**

AI methods that use transfer learning can cope with arbitrary data, since the inherent knowledge can be transferred to an AI through transfer learning methods. In addition, such AI methods can generalise better and require less learning data. Due to the additional development effort, the transfer learning AI is smarter and can therefore recognise the synthetic vehicle in this example with the help of other **synthetic data (triangle).**

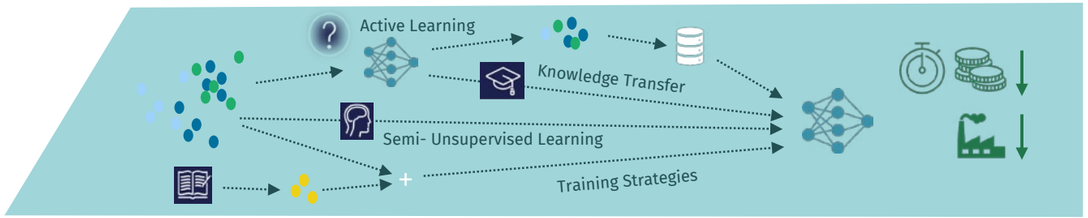
Transfer learning in a nutshell (©Bosch | Gemeinfrei)

Didactics

Didactics enables learning, by reflecting and structuring the learning process. While in principle it is clear how to train a neural network with annotated data, the acquisition of the data is a time and resource consuming process, as human annotators are required. Improving this process to make the learning easier for deep neural networks is the task of the project area didactics. In particular, in the presence of various deltas, it would be inefficient to start learning from scratch over and over again. Therefore, in this project area we develop new approaches to learn with only a few data points annotated by humans or with no such labels at all (semi- and unsupervised learning), to accelerate learning by optimizing network architectures with regard to training or to improve the learning algorithms (training organization), to learn with less data by selecting the most informative samples (active learning), or to enable learning by building on prior successful learning processes (knowledge transfer). The project area didactics provides the common house to foster research in all these directions and has resulted in a huge number of remarkable methodological results and scientific publications.



The process of supervised learning. Data is collected and labeled by human annotators, before it can be used for training of neural networks. This process oftentimes is expensive and consumes a lot of energy for training specialized neural networks on various tasks.



Didactics searches for better solutions. Only a few data points are annotated according to special acquisition strategies, networks reuse prior knowledge from related tasks or distill knowledge from teacher networks. In addition, with augmentation the dataset can be enhanced. Thereby, cost is reduced and training efficiency is enhanced.

Automotive Suitability

Motivation

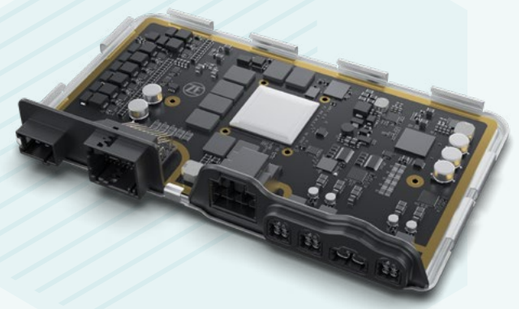
The process commonly used in the industry to engineer automotive AI systems is to develop, train, and verify AI functions in the lab using recorded data. Only the finished AI system is transferred to the vehicles during production or an update. Two problems arise here: First, the high-performance computer hardware in the lab and the embedded hardware in the vehicle differ significantly, and second, the situations in which a vehicle encounters in the real world may differ significantly from the previously recorded training and test data.

Embedded Systems

The problems that arise in the transition to automotive embedded hardware are the limited resources. Both the computing speed and the available working memory are usually much lower in such systems than in usual laboratory computers. Despite these limitations, the AI system in the vehicle should nevertheless work reliably within the time limits specified for vehicle safety and demonstrate comparable performance to the laboratory system. To bridge this embedded systems delta, various techniques were developed in the project to reduce the resource requirements of an AI in the vehicle without severely compromising its function.

Real World

The problems that arise when using a system trained and tested on no matter how much pre-recorded or generated data are more multifaceted. As the environment is constantly changing, novel objects will always appear thatwere not known at the time the AI system was created. In addition, there is an uncontrollably large amount of very rare and strange objects, behaviors, and conditions that cannot be fully represented in any data set. To bridge this real world delta, methods were developed in the project to increase the robustness of AI systems even in the presence of unexpected or unknown scenarios. New efforts in the design of automotive AI systems foresee a continuous monitoring of the AI functions in the field. Inthis way, data can be collected from the fleet that will lead to further design iteration to improve the response of the AI system to unexpected and unknown situations.



ZF's ProAI offers highest compute ressorces for automotive embedded AI

Environment

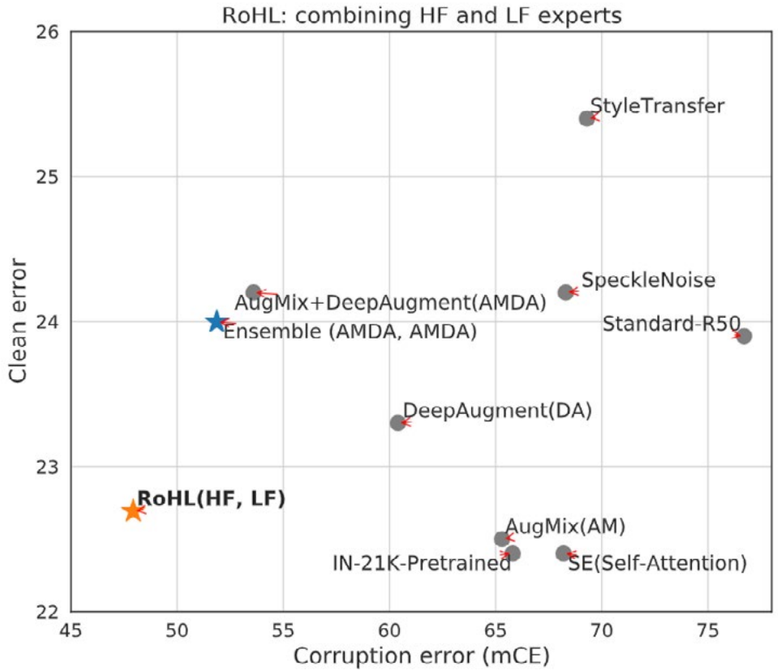
This category stands for changes in the surrounding over time. It summarizes long-term changes, such as those in weather and season, short-time changes as day and night as well as deltas between different countries and traffic conditions like traffic density or new and unfamiliar objects.

Improving robustness against common corruptions with frequency biased models 24
Introducing Intermediate Domains for Effective Self-Training during Test-time 26
Robustness Against Noisy Labels Through Uncertainty Estimation for LiDAR-based Semantic Segmentation 28
An Unsupervised Domain Adaptive Approach for Multimodal 2D Object Detection in Adverse Weather Conditions. 30
A Low-Complexity Approach for Domain Adaptation 32
Continual Learning for Model-Based Reinforcement Learning 34
Motion Capture-based Virtual Reality Co-Simulation. 36
Domain Shift Quantification using Activations 38
SceneNeRF: 3D Reconstruction of Real-World Scenes 40
Environmental adaptation and self-attention in the context of unsupervised domain adaptation 42
Detection of critical weather situations in scenario-based traffic simulations using optimization techniques 44

Improving robustness against common corruptions with frequency biased models

Tonmoy Saikia, Thomas Brox, University of Freiburg | **Cordelia Schmidt**, INRIA

CNNs perform remarkably well when the training and test distributions are i.i.d, but unseen distortions, such as weather distortions can cause large drop in performance. Image corruption types have different characteristics in the frequency spectrum and would benefit from a targeted type of data augmentation, which, however, is often unknown during training. We introduce a mixture of two expert models specializing in high and low-frequency robustness, respectively. Moreover, we propose a new regularization scheme that minimizes the total variation (TV) of convolution feature-maps to increase high-frequency robustness. The approach improves on corrupted images without degrading in-distribution performance. We demonstrate this on ImageNet-C and on an automotive dataset, both for object classification and object detection.

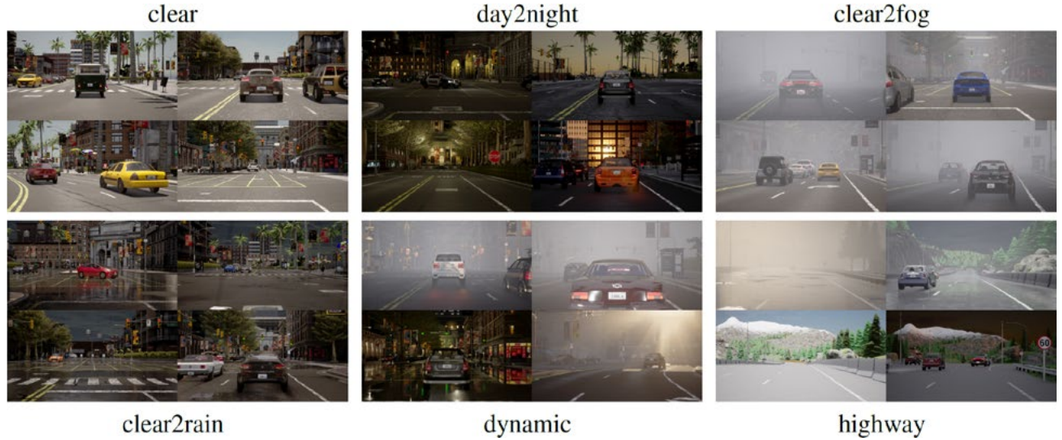


Trade-off between in-distribution performance (clean error) and out-of-distribution robustness (Corruption error). The proposed mixture of a low-frequency and a high-frequency expert shows a very good trade-off and yields the highest robustness. (© University of Freiburg)

Introducing Intermediate Domains for Effective Self-Training during Test-time

Robert Marsden, Mario Döbler, Bin Yang, University of Stuttgart

Experiencing domain shifts during test-time is nearly inevitable in practice and likely results in a severe performance degradation. To overcome this issue, test-time adaptation continues to update the initial source model during deployment. A promising direction are methods based on self-training which have been shown to be especially well suited for gradual domain shifts, since reliable pseudo-labels can be provided. While many domain shifts in reality evolve gradually, this does not always hold. Therefore, we aim to create an artificial intermediate domain during test-time which divides the current domain shift into a more gradual one, enabling to perform effective self-training. To investigate gradual shifts in the context of urban scene segmentation, we publish a new benchmark: CarlaTTA. It enables the exploration of several non-stationary domain shifts

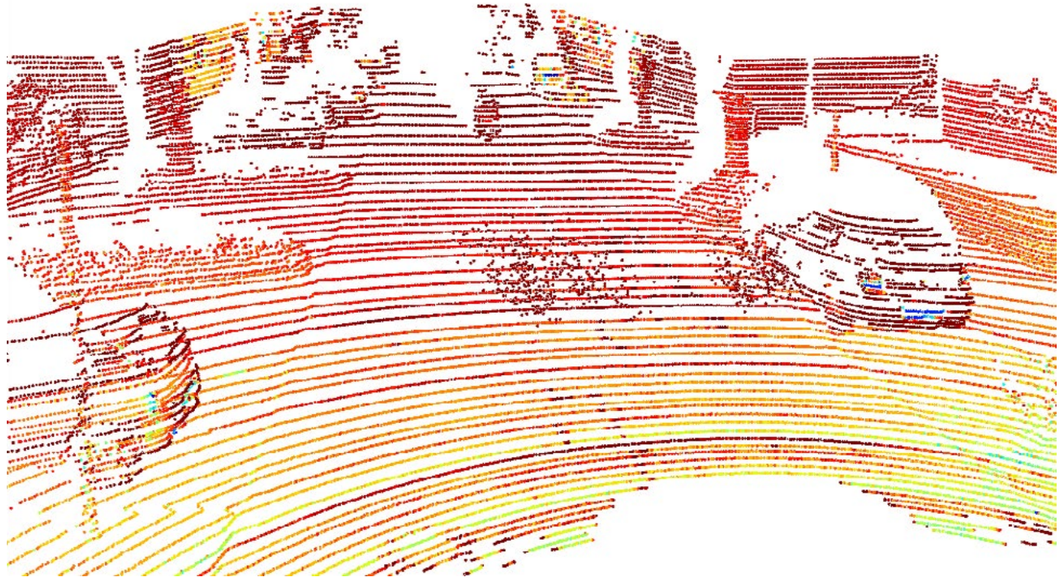


A few example images for each split contained in the new benchmark CarlaTTA. Starting from the source domain „clear“, every domain evolves gradually over time. The split „dynamic“ combines multiple domain shifts at a time and thereby even creates new ones. „Highway“ further introduces a shift in the class priors. (© University of Stuttgart)

Robustness Against Noisy Labels Through Uncertainty Estimation for LiDAR-based Semantic Segmentation

Mariella Dreissig, Florian Piewak, Andras Tuezkoe, Mercedes-Benz Cars AG

The predictive performance of any deep learning-based environment perception model for autonomous driving is partially governed by the quality of the underlying dataset. Systematic problems with the dataset and the respective labels can have a huge impact on the internal representation of the feature landscape the model infers from the ingested data. We discovered that state-of-the-art uncertainty estimation methods provide a basis for identifying and dealing with problematic label definitions. We furthermore developed a lean method on robustness against noisy labels using an hierarchical abstraction loss. We suggest that it can be applied to different domain shifts present in the data.



Noisy LiDAR measurements through fog. Fog-augmented [1] point cloud from the KITTI dataset [2].

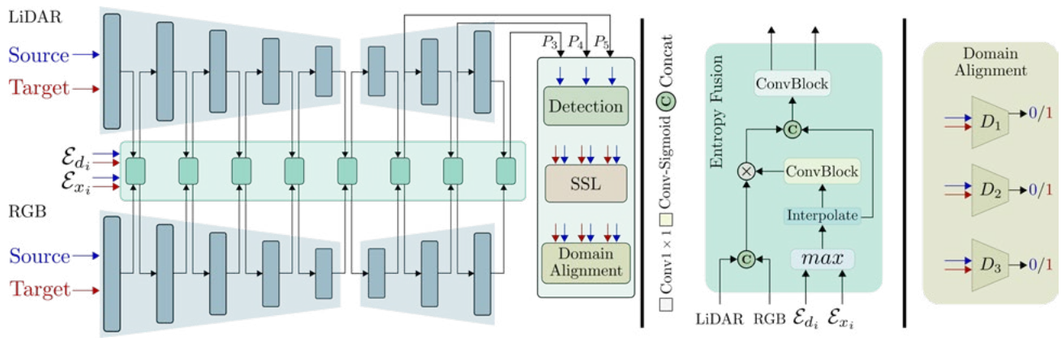
[1] M. Hahner, et al., “Fog simulation on real lidar point clouds for 3D object detection in adverse weather,” in IEEE ICCV, 2021.

[2] A. Geiger, P. Lenz, and R. Urtasun, “Are we Ready for Autonomous Driving? The KITTI Vision Benchmark Suite,” in IEEE CVPR, 2012.

An Unsupervised Domain Adaptive Approach for Multimodal 2D Object Detection in Adverse Weather Conditions

George Eskandar, Robert Marsden, Pavithran Pandiyan, Mario Döbler, Bin Yang, University of Stuttgart

While deep learning architectures that fuse vision and range data for 2D object detection have thrived in recent years, the corresponding modalities can degrade in adverse weather conditions, leading to a performance drop. Although domain adaptation methods attempt to bridge the domain gap between source and target domains, they do not extend to hete-rogeneous data distributions. We propose an unsupervised domain adaptation framework, which adapts a 2D object detector for RGB and LiDAR sensors to a target domain featuring adverse weather conditions. Experiments performed on the DENSE dataset show that our method outperforms state-of-the-art unimodal methods in the single-target and multi-target domain adaptation settings.

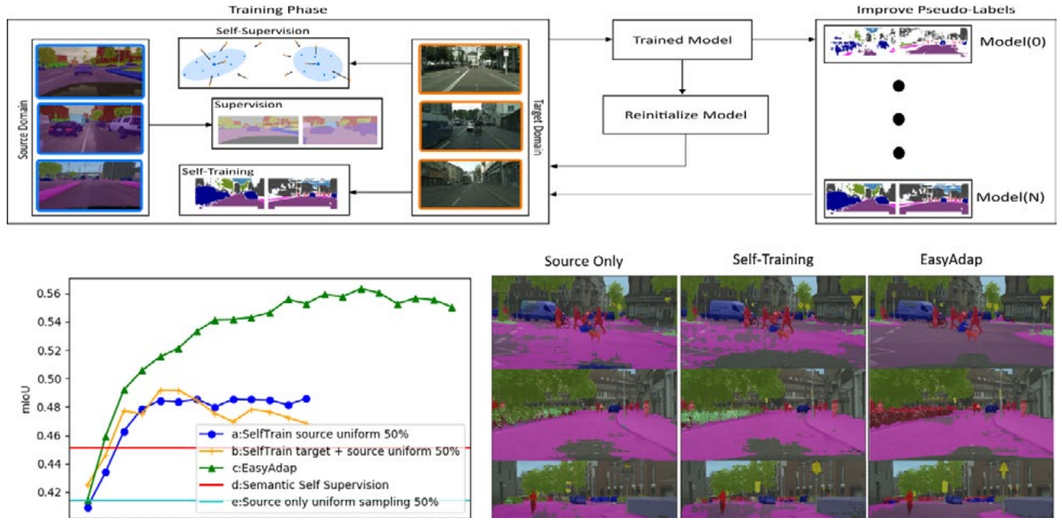


The proposed approach consists of three components: a data augmentation scheme to simulate weather distortions, a cross-domain foreground object alignment, and a subnetwork to learn pretext tasks in a self-supervised way.

A Low-Complexity Approach for Domain Adaptation

Joshua Niemeijer, Jörg P. Schäfer, Deutsches Zentrum für Luft- und Raumfahrt e.V.

We present a low-complexity approach to address the domain discrepancy. We aim to align both distributions through semantic Self-Supervision. To that, we compute the class prototypes that represent the classes in the feature space. We assume that target domain feature representations are closer to the correct class centroids than to the incorrect ones. We exploit this property to improve this clustering and increase the target domain's segmentation quality. Given the improvements on the target domain, we can generate high-quality pseudo labels for self-training on the target domain. Self-training on the target domain results in the feature space's alignment of target and source domain distributions. This aids semantic clustering. Thus we end up with the synergistic effect seen in Figure 3 by iteratively improving the pseudo labels.

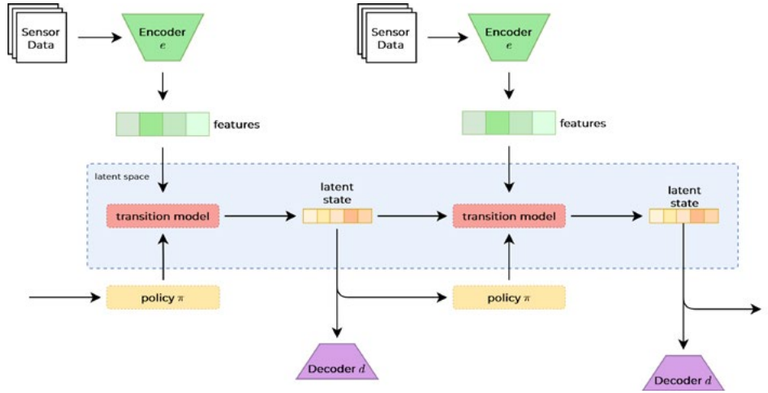


Top: An overview of the internal dependencies of our iterative approach for unsupervised domain adaptation. Bottom left: Segmentation quality over the iterations (green: the developed method). Bottom right: Results of the unsupervised domain adaptation approach for a adaptation from synthetic (GTA5) to real (Cityscapes) data.

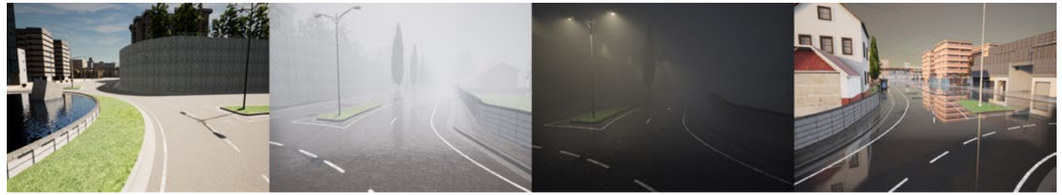
Continual Learning for Model-Based Reinforcement Learning

Tim Joseph, FZI Forschungszentrum Informatik, Karlsruhe

Keeping all collected data indefinitely for training is often infeasible. However, training a model in a naive way whenever new data is available leads to catastrophic forgetting, a phenomenon that describes the abrupt loss of knowledge of previously learned tasks as information relevant to the current task is incorporated. We use a model-based reinforcement learning agent and regularize its components to fight forgetting between different tasks. For example, one task is to drive on a sunny day while another task may be to drive in a rainy night. The agent should adapt to both scenarios after having seen them. Also, after having experienced and learned on the second task, performance of the first task should improve. Our work shows, that simply applying continual regularization methods is not sufficient to succeed in the overall control task.



A general overview of our used architecture. Our agent consists of four modules that are regularized independently to combat forgetting.



Different weather settings in CARLA in which we train and evaluate our agent.

Motion Capture-based Virtual Reality Co-Simulation

Markus Rehmann, Michael Brunner, Cristóbal Curio, Reutlingen University

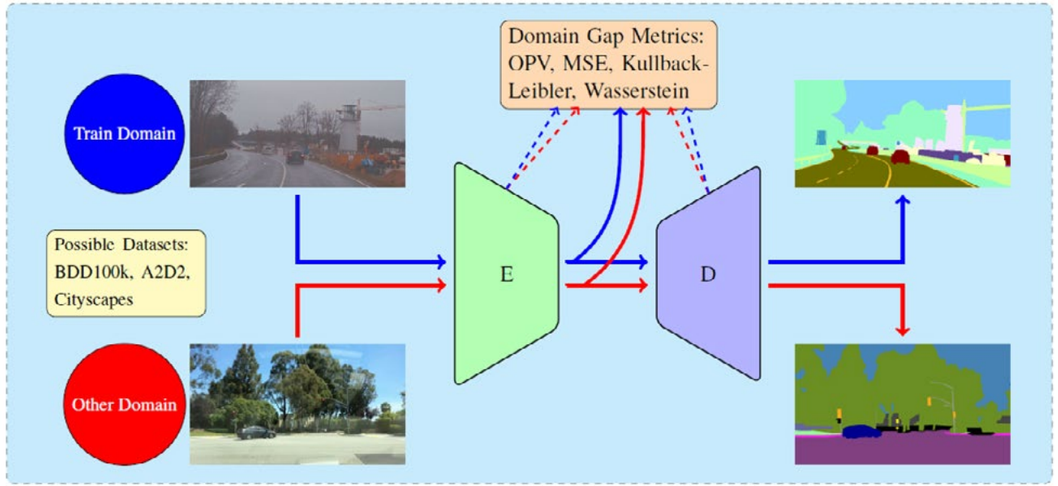
Many current simulation environments suffer from limited and repetitive human animations. Using these unrealistic animations for model training may cause wrong predictions and reduced accuracy of human behavior models on more realistic data. By combining motion capture and Virtual Reality, scene-relevant animations can be created in which the actor perceives and can interact with the virtual environment. Obtaining realistic human behavior models, which support autonomous systems in better understanding humans, largely benefit from scene-relevant interactions during the recording process.

Co-Simulation outside view (left), Virtual Reality view (right) (© Markus Rehmann, Reutlingen University)

Domain Shift Quantification using Activations

Manuel Schwonberg, Indrani Sarkar, Nico Schmidt, CARIAD SE

Unsupervised Domain Adaptation (UDA) aims to adapt Deep Neural Networks (DNNs) to new domains by only accessing unlabeled target data. The proposed approaches focus on increasing the performance scores like the mIOU to report the increased adaptation capability on the target domain. Little to no knowledge exist about the internal behavior and mechanism within the DNNs under domain shift. We propose to utilize distribution distances like Wasserstein or Frechet Inception Distance (FID) to quantify the domain shift between two domains in an unsupervised manner by only accessing their network activations. We find that the layers of the network are differently strong affected by the domain shift and that our metrics are not directly correlated with the mIoU.

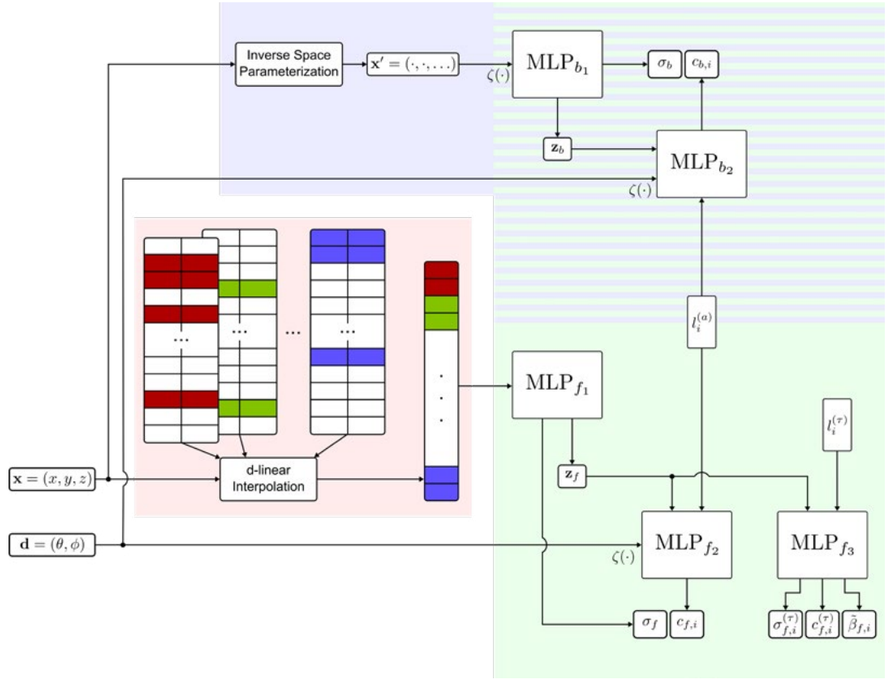


Domain Shift Quantification using Activations (© CARIAD SE)

SceneNeRF:
3D Reconstruction of Real-World Scenes

Thies de Graaff, Deutsches Zentrum für Luft- und Raumfahrt e.V.

The generation of synthetic data samples for AI training and testing, as well as scenario-based testing of entire automated driving functions requires a great variety of realistic virtual worlds. This work investigated the automated reconstruction of real-world scenes based on images and sparse point clouds, gathered from sensors attached to a vehicle. Our approach is based on Neural Radiance Fields (NeRF) and incorporates different ideas into an overall architecture to be able to cope with large traffic scenes that involve highly-dynamic traffic. Current results build a good foundation to be refined manually, reducing the overall human effort in creating virtual worlds.

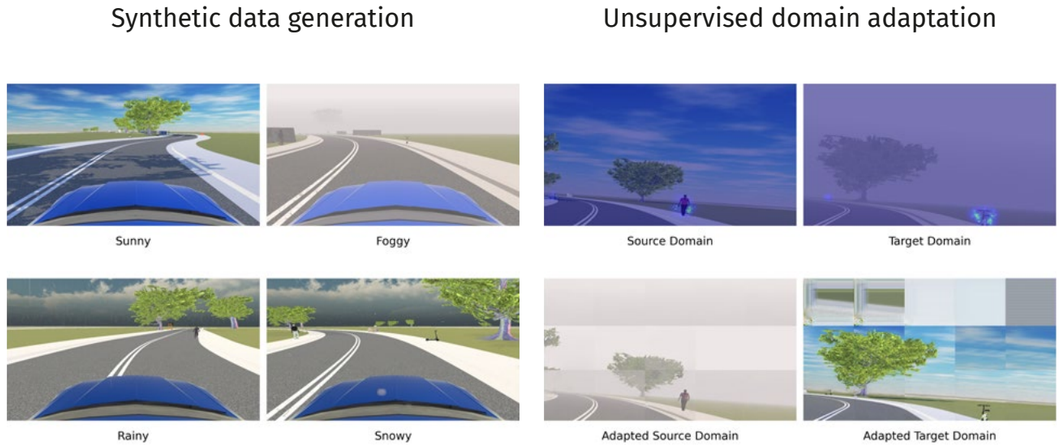


Overall SceneNeRF architecture to reconstruct 3D scenes from real-world image data (© DLR)

Environmental adaptation and self-attention in the context of unsupervised domain adaptation

Vinu Vijayakumaran Nair, Markus Rehmann, Michael Brunner, Cristóbal Curio, Reutlingen University

Simulations are ideal tools for generating large amounts of training data of challenging weather conditions, including novel mobility classes such as e-scooters. The effectiveness of using simulated data for environmental adaptation with unsupervised domain adaptation needs further investigation. The use of self-attention methods can improve model performance in the real domain, but in combination with unsupervised domain adaptation (UDA), it is possible to further improve performance. We investigated the environmental adaptation of an object detection model and the transferability of attention-based pose estimation models in the context of UDA on different datasets.

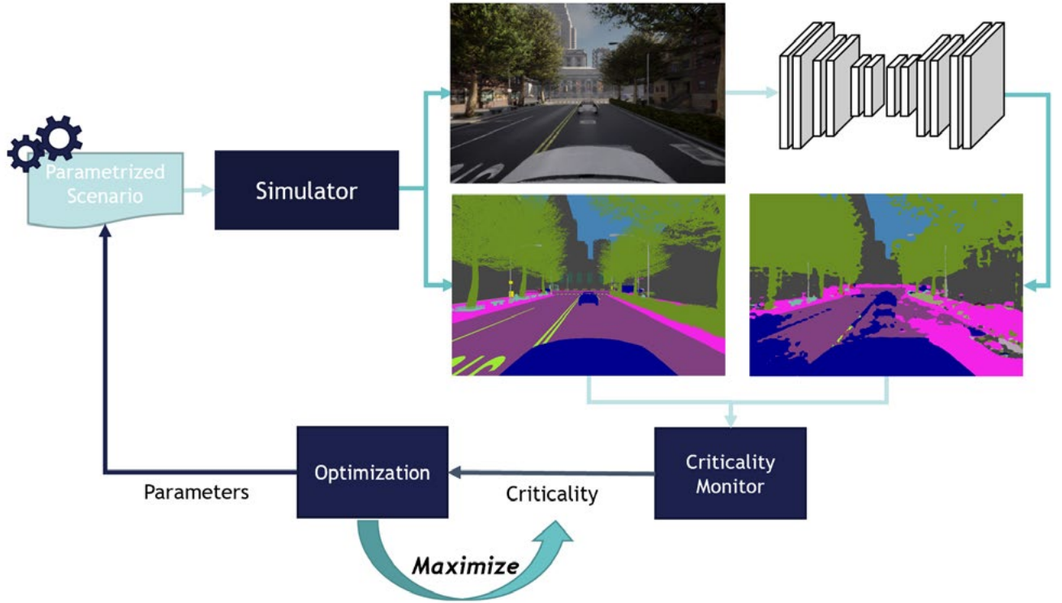


Highlights and overview of our approach, Synthetic data generation for various environmental conditions, Unsupervised domain adaptation for object detection and pose estimation tasks (© Reutlingen University)

Detection of critical weather situations in scenario-based traffic simulations using optimization techniques

Daniel Grujic, Thies de Graaff, Günter Ehmen, Möhlmann, Deutsches Zentrum für Luft- und Raumfahrt e.V.

The performance of the neural network depends heavily on the training dataset. Since these training data sets usually can not cover all possible situations, there is always a risk to face a critical situation in the target domain where the detection rate of the network is too low. Therefore, we use an optimization-based approach to automatically identify critical simulation parameters that can lead to these critical situations. We apply this approach to the use-case “weather”. Here we want to find the set of weather parameters for a given scenario, where the neural network - the system-under-test - performs worst.



Sketch of the scenario-based optimization workflow. (© Deutsches Zentrum für Luft- und Raumfahrt e.V.)

Sensors

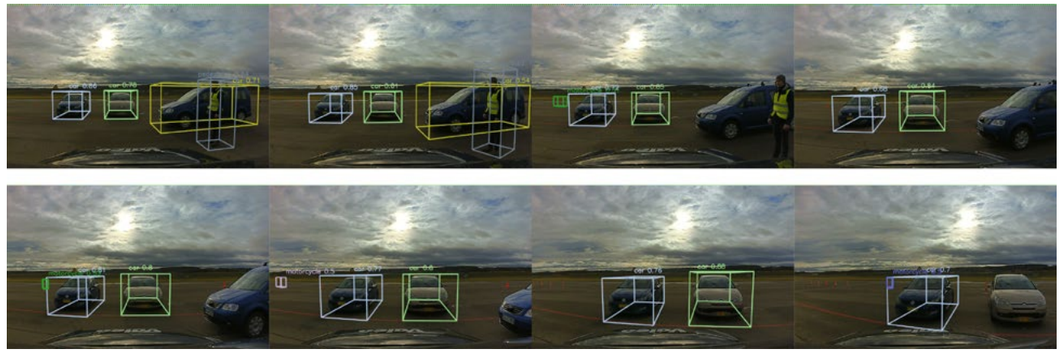
Sensors, as Lidar, Radar or Camera are used in transfer learning to identify objects in autonomous driving such as other vehicles, pedestrians, and traffic signs in the environment.

3D Detection and Tracking From LiDAR Point Clouds As a Pre-Processing Step for Active Learning 48
Processing of vehicle sensor data 50
Real Data Acquisition with Ground Truth 52
Auxiliary Task-Guided CycleGAN for Black-Box Model Domain Adaptation 54
Bridging Domain Gaps in Lidar Perception 56
Lidar Upsampling with Sliced Wasserstein Distance 58
TransFuser: Imitation with Transformer-Based Sensor Fusion 60
HALS: A Height-aware Lidar Super-Resolution Approach for Autonomous Driving 62

3D Detection and Tracking From LiDAR Point Clouds As a Pre-Processing Step for Active Learning

Florian Bogner, Norman Müller, Technical University of Munich

We base our detection and tracking method on Centerpoint, which we successfully adapt to the project's dataset. As a result, we are able to auto-label 3D bounding boxes and achieve consistent inter-frame tracking. As each detected instance includes a confidence, we can derive a metric for overall frame confidence to be used as an input to subsequent Active Learning. Especially regarding the fact that we exclusively worked with nuScenes training data, the qualitative results (depicted on the right) we were able to achieve on unlabeled project data are very promising.

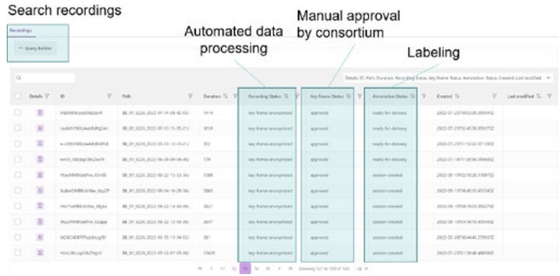


Qualitative evaluation of detection and tracking on project data (© TUM)

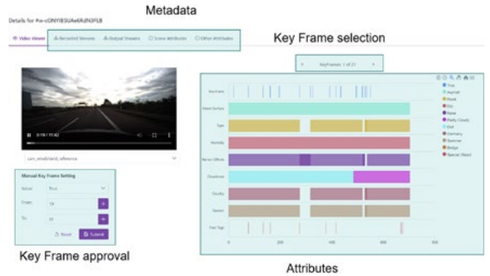
Processing of vehicle sensor data

Dennis Neumann, Daniel Ruf, Roshan Muthaiya, Christian König, Luxoft GmbH

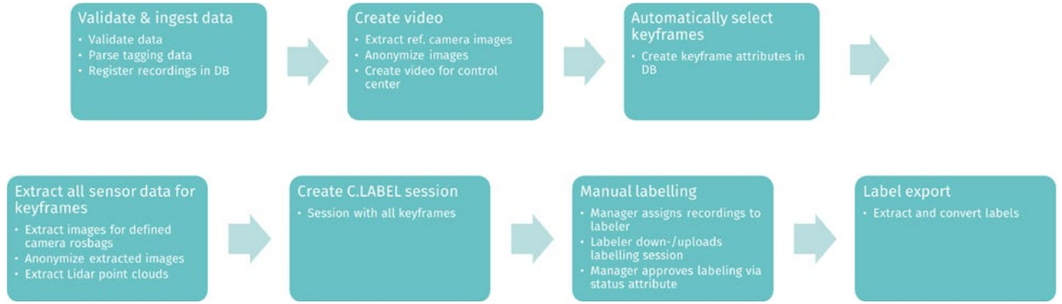
To be useful for training machine learning models, sensor data from test drives needs to be processed and labelled. In addition, privacy legislation demands all video data to be anonymized before further processing. This processing comprises a user interface where recordings can be viewed and frames for labelling can be selected. Here, metadata for each recording is shown along with the status of processing. Furthermore, a pipeline of sequential automatic operations is implemented, which consists of data ingestion, video creation and anonymization, automatic selection of key frames and extraction of data for labelling.



Overview of the recording metadata (© Luxoft GmbH)



User interface for the keyframe approval (© Luxoft GmbH)



Pipeline for automatic processing of sensor data (© Luxoft GmbH)

Real Data Acquisition with Ground Truth

Franz Andert, Joshua Niemeijer, Jörg Schäfer, Deutsches Zentrum für Luft- und Raumfahrt e.V.
Silas Maile, DXC Luxoft | **Tobias Wagner, Christian Witt**, Valeo

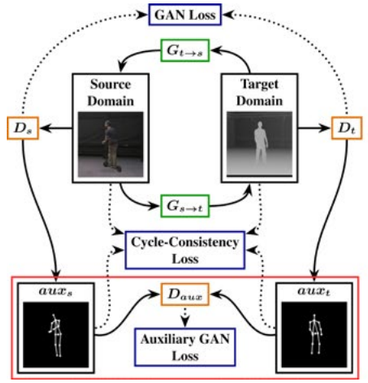
To close existing gaps in publicly available datasets, KI Delta Learning created a complete new set of real data with multiple sensors and in different conditions. A Mercedes-Benz V-class van was equipped with a variety of cameras, LIDARs, RADARs, and other hardware. During the project, the car was driven in rural and urban areas in Germany and Italy. In Berlin and Braunschweig, measurements from a second DLR vehicle and from the stationary DLR test fields were included, and special maneuvers with mutual interaction of both cars were performed. We recorded 193 hours of data and about 7,000,000 sensor frames. 18,000 image and LiDAR samples were labeled for semantic and instance segmentation and 3D bounding box detection.

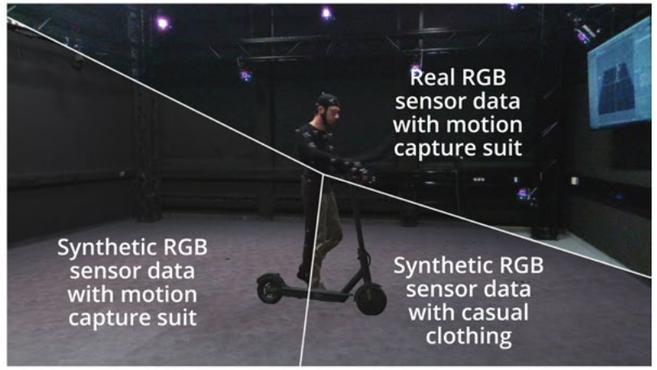
Overtaking maneuver at Test Field Lower Saxony. Data Acquisition with two cars, and with object reference from road site cameras (© DLR)

Auxiliary Task-Guided CycleGAN for Black-Box Model Domain Adaptation

Michael Brunner, Markus Rehmann, Cristóbal Curio, Reutlingen University

Usually, existing DA methods are targeted at specific tasks and require access to the source model which is a major drawback when only a black-box model is available. We implemented a CycleGAN-based approach suitable for black-box source models. An auxiliary task is used to support the transfer of task-related information across domains. We have shown the effectiveness for the challenging task of 2D human pose estimation and compared our results in four different domain adaptation settings to CycleGAN and RegDA, a state-of-the-art method for unsupervised domain adaptation for keypoint detection.



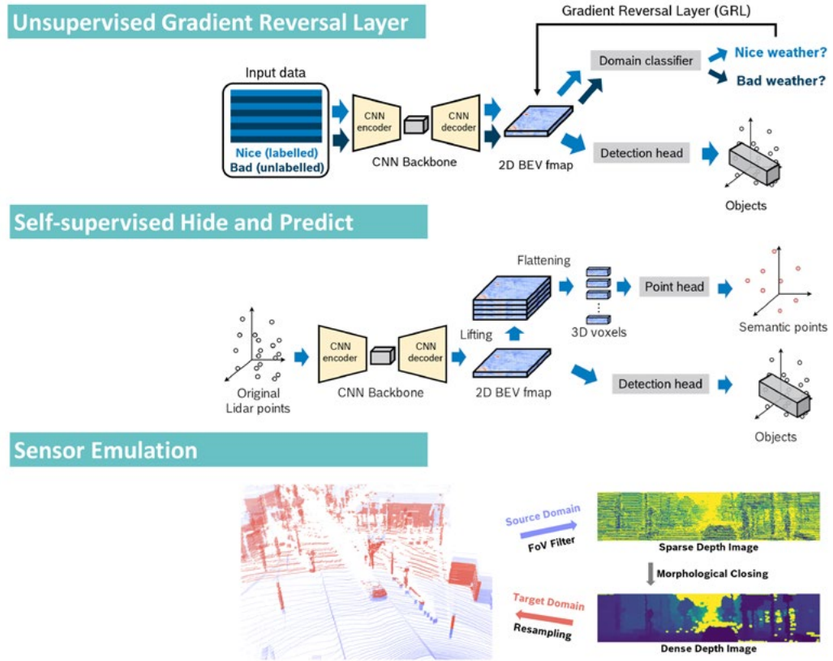


Left: Our method extends CycleGAN with additional auxiliary GAN and cycle consistency losses (highlighted red box). Right: A sample frame of our paired dataset showing three different domains. (M. Essich, M. Rehmann, and C. Curio, “Auxiliary Task-Guided CycleGAN for Black-Box Model Domain Adaptation,” in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Jan. 2023, pp. 541–550.)

Bridging Domain Gaps in Lidar Perception

Jasmine Richter, Florian Faion, Di Feng, Thomas Nürnberg, Claudius Gläser, Robert Bosch GmbH

We explored three different approaches to reduce the delta sensor and delta weather domain gap in Lidar 3D object detection where some modify the features and others the input data. We showed that all of them have the potential to reduce the gap to a certain degree (5-15% relative performance gain in the target domain), but none was able to completely close it. In addition to bridging gaps, all approaches (and combinations of them) can be used as general-purpose augmentation methods to increase robustness of detectors.



Overview of the considered approaches to reduce the domain gap in Lidar 3D object detection. (© Bosch)

Lidar Upsampling with Sliced Wasserstein Distance

Artem Savkin, Sebastian Wirkert, BMW AG
Yida Wang, Nassir Navab, Federico Tombari, Technical University of Munich

Data-driven perception systems require data acquisition and annotation on scale which is an expensive and inflexible process. We address the problem of sensor-tosensor domain adaptation to avoid re-acquiring or re-annotating the data and focus on the sensor setup with low- and high-resolution lidars and the task of upsampling. Contrary to recent methods, the proposed technique demonstrated the ability to reconstruct fine scan patterns of lidar point clouds. It also showed improved lidar upsampling performance according to established metrics.

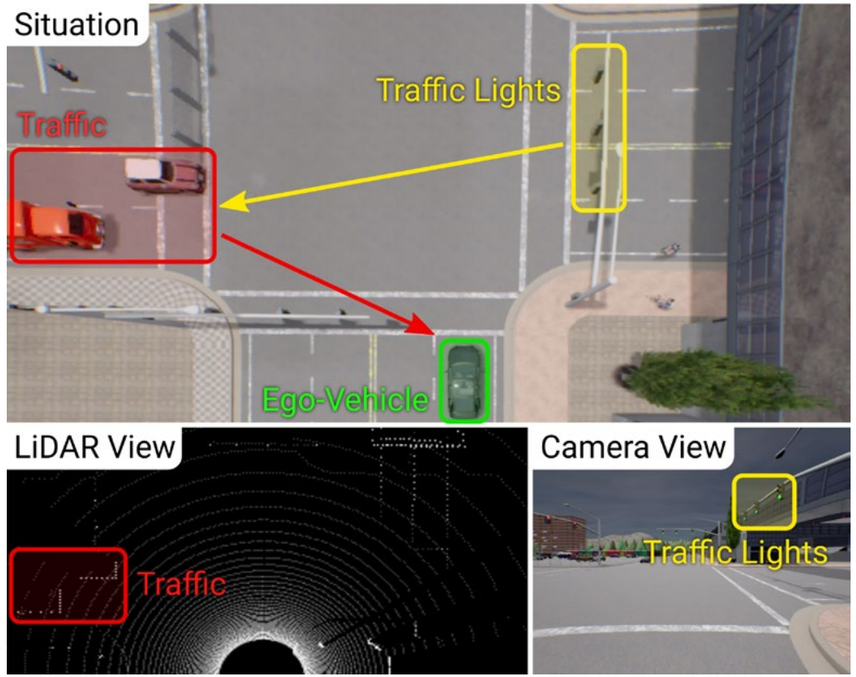
This work was published in the IEEE Robotics and Automation Letters

Lidar upsampling results (middle) on KITTI val patches with input samples (top) and ground truth (bottom). (© IEEE)

TransFuser: Imitation with Transformer-Based Sensor Fusion

Kashyap Chitta, Bernhard Jaeger, Zehao Yu, Katrin Renz, Andreas Geiger, University of Tübingen
Aditya Prakash, University of Illinois Urbana-Champaign

How should we integrate representations from complementary sensors for autonomous driving? We propose TransFuser, a mechanism to integrate image and LiDAR representations using self-attention. Our approach uses transformer modules at multiple resolutions to fuse perspective view and bird's eye view feature maps. We experimentally validate its efficacy on a challenging new benchmark with long routes and dense traffic, as well as the official leaderboard of the CARLA urban driving simulator. Compared to geometry-based fusion, TransFuser reduces the average collisions per kilometer by 48%.

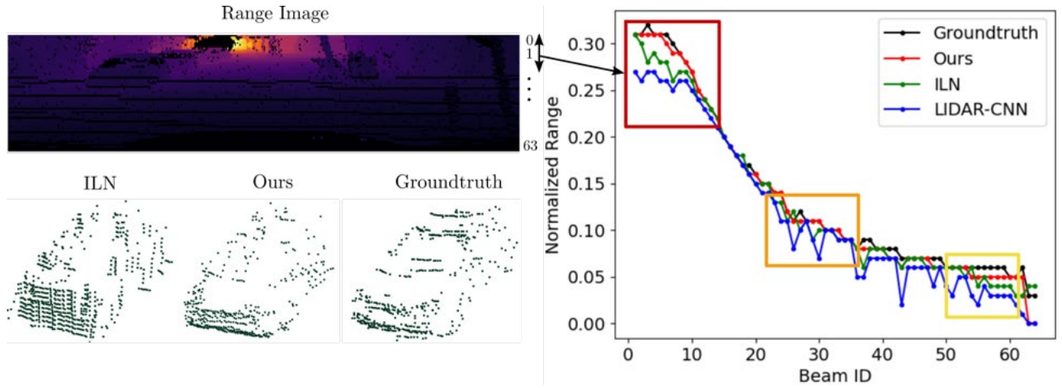


To safely navigate, an agent must understand the interaction between geometrically distant elements of the scene. Our TransFuser model integrates geometric and semantic information across multiple modalities via attention mechanisms to capture global context, leading to safe driving behavior in CARLA. (© University of Tübingen)

HALS: A Height-aware Lidar Super-Resolution Approach for Autonomous Driving

George Eskandar, Sanjeev Sudarsan, Bin Yang, University of Stuttgart

Upsampling lidar pointclouds is a promising approach to gain the benefits of high resolution while maintaining an affordable cost. We introduce a novel lidar upsampling model, HALS: Height-Aware Lidar Super-resolution. We exploit the observation that lidar scans exhibit a height-aware range distribution and adopt a generator architecture with multiple upsampling branches of different receptive fields. The branches' outputs are fused using confidence maps to model the branches' uncertainty. HALS regresses polar coordinates instead of spherical coordinates and uses a surface-normal loss for the first time in the training pipeline of lidar upsampling. Extensive experiments show that we achieve state-of-the-art performance on 3 real-world lidar datasets.



When projected on a 2D spherical range image, a lidar scan exhibits a height-dependent range distribution. We find that far away objects (high range values) are usually represented in the upper part of the range image (Beam ID 0 corresponds to the highest row in the range image). Our model can better follow this distribution than previous approaches. Extracted cars from upsampled lidar scans demonstrate that the overall geometry and shape of foreground objects are better preserved.

Active Learning

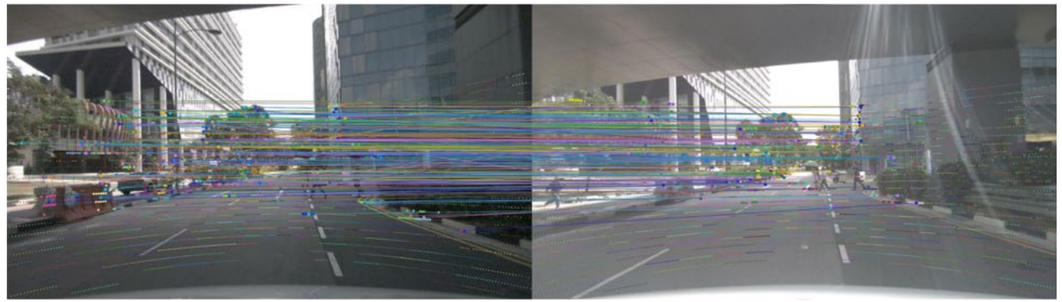
We acquire only a few, but the most important examples from a large stream of data for labeling, in order to improve the perception in automated driving.

Active Learning On Dynamic Scenes Using Multi-View Consistency 66
Active Learning based on a Taxonomy for Scene Description 68
Active learning for semantic segmentation in realistic driving scenarios 70
Consistency-based Active Learning for Semantic Segmentation 72

Active Learning On Dynamic Scenes Using Multi-View Consistency

Daniel Derkacz-Bogner, Norman Müller, Technical University of Munich

We extend an active learning pipeline using multi-view consistency to the domain of dynamic LIDAR sequences. Due to the missing correspondence between points in different timesteps we evaluate two point matching approaches, a nearest neighbor matching which only works on static scene parts and a 2D SIFT-LIDAR matching which is able to correspond static as well as dynamic objects. Our method on static scenes achieves 70% of the baseline performance while only using 20% of the data. Due to the limited amount of dynamic point matches, we could not achieve similar results on the dynamic scene setting.

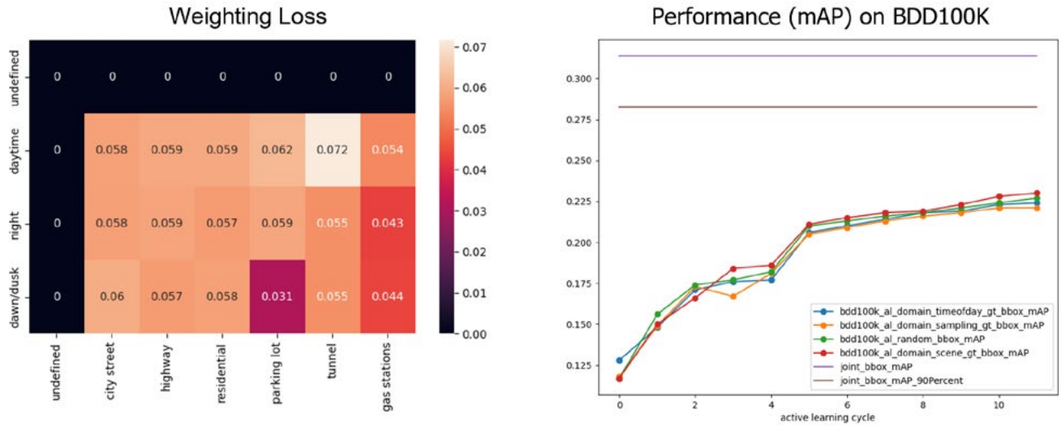


Visualization of our 2D SIFT-LIDAR matching (©TUM)

Active Learning based on a Taxonomy for Scene Description

Christian Witte, Syed Saqib Bukhari, Georg Schneider, ZF Group

Active Learning describes an iterative training paradigm that selects data samples to be annotated. We derive a taxonomy for describing an automotive scene and leverage domain information for the sampling process. This allows the Active Learning process to evaluate the network's performance for particular scene descriptions and to sample challenging scenarios. Further, we introduce a weighting loss that also incorporates domain distributions. Our experiments show that the selection solely based on domain information, even with the modified loss, yields no advantage over the random sampling baseline.

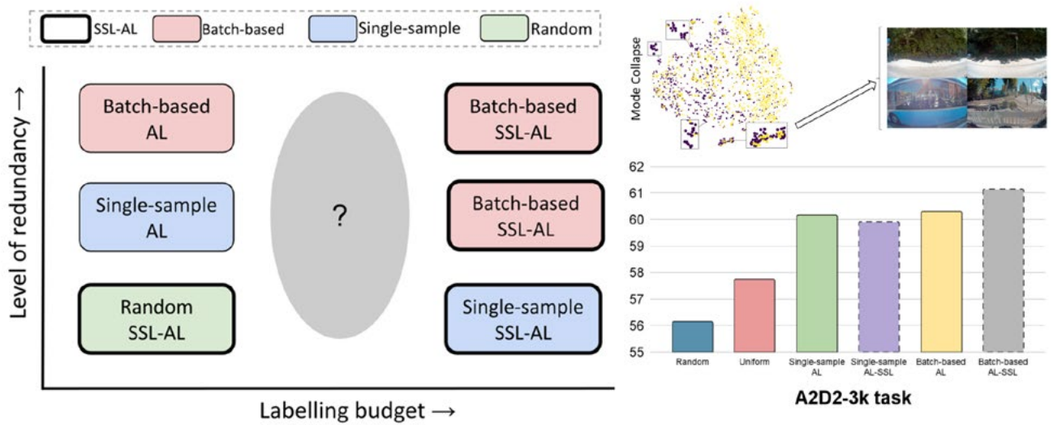


Weighting Loss for Query Selection for Time of Day and Scene Domains (left), Active Learning Performance of a Object Detection Network for Single and Joint Domains (right) (©ZF Group)

Active learning for semantic segmentation in realistic driving scenarios

Joshua Niemeijer*, Jörg P. Schäfer, Deutsches Zentrum für Luft- und Raumfahrt e.V.
Sudhanshu Mittal*, Thomas Brox, University of Freiburg | * Indicates equal contribution

We show that the data distribution is decisive for the performance of the various active learning objectives proposed in the literature. Particularly, redundancy in the data, as it appears in most driving scenarios and video datasets, plays a significant role. We demonstrate that integrating semi-supervised learning with active learning can improve performance when the two objectives are aligned. Our experimental study shows that current active learning benchmarks for segmentation in driving scenarios are unrealistic since they operate on data already curated for maximum diversity. Accordingly, we propose a more realistic evaluation scheme in which the value of active learning becomes clearly visible, both by itself and in combination with semi-supervised learning.

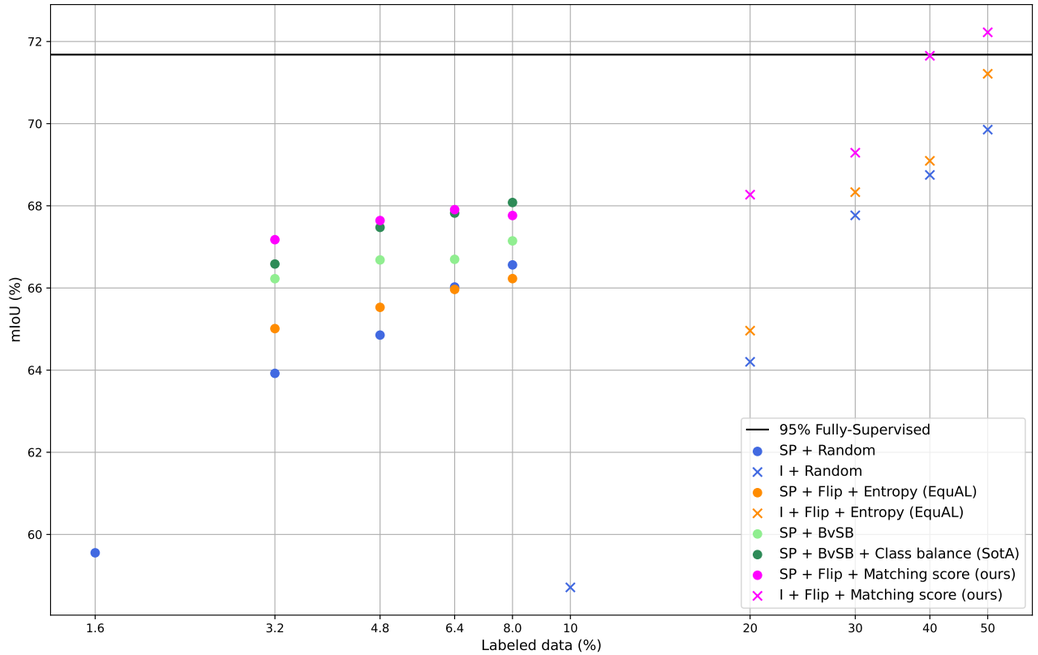


Left: Overview of how to chose the acquisition function and whether to apply semi supervised learning w.r.t. the redundancy of the dataset and the labeling budget in each active learning circle. Top Right: Illustration of the Mode Collapse problem on A2D2 data. Bottom Right: Results on our newly proposed realistic Benchmark for Active Learning.

Consistency-based Active Learning for Semantic Segmentation

Stefan Matthes, fortiss GmbH | **Sebastian Wirkert**, BMW AG

We investigated active learning (AL) approaches that select data for labeling based on the consistency of the predictions of the model being trained. We propose to predict the semantic segmentations on the original and flipped images and rank by the number of mismatched pixels. We compare our method with uncertainty-based approaches under two regimes where either entire images or image regions are queried for labeling. As shown in the right figure, our proposed method performs equally well or better than previous AL approaches (measured in mIoU) in both acquisition regimes, yet is simple to implement.



Comparison of consistency- and uncertainty-based AL approaches on Cityscapes for image- (I) and region-based (SP=superpixel) semantic segmentation. We randomly select the first 10% of images (or 1.6% of regions) to train our initial model and query the same amount per round according to the respective acquisition function. (© fortiss | BMW)

Knowledge Transfer

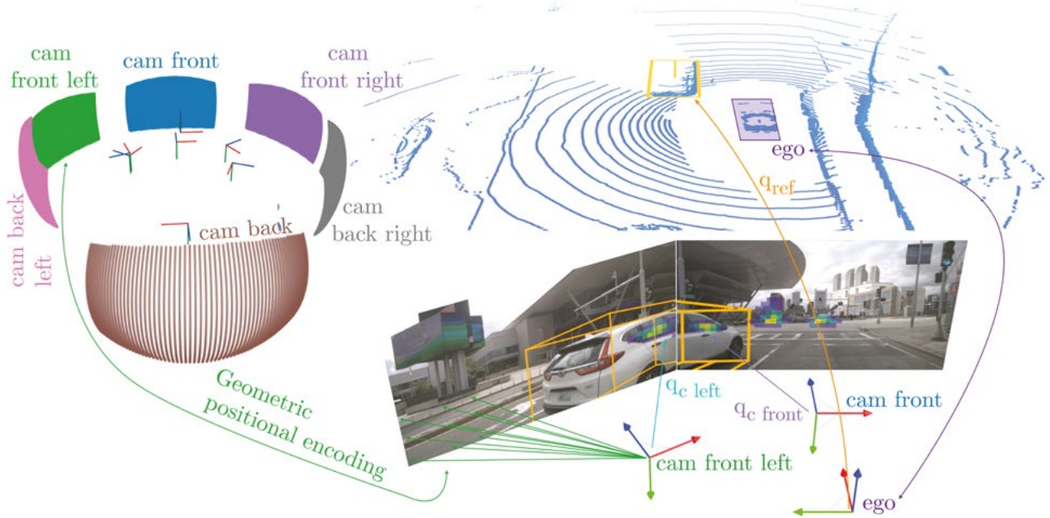
The basic question in Knowledge Transfer is, how knowledge already learned by one network can be transfered to another network.

SpatialDETR: 3D Object Detection from Multi-View Camera Images with Global Cross-Sensor Attention 76
Knowledge Transfer for Multitask and Downstream Tasks 78
Domain Generalization and (Continuous) Unsupervised Domain Adaptation 80
USIS: Unsupervised Semantic Image Synthesis 82
CRAT-Pred: Vehicle Trajectory Prediction with Crystal Graph Convolutional Neural Networks and Multi-Head Self-Attention 84

SpatialDETR: 3D Object Detection from Multi-View Camera Images with Global Cross-Sensor Attention

Simon Doll, Richard Schulz, Lukas Schneider, Mercedes-Benz AG | **Hendrik P.A. Lensch**, University of Tübingen
Markus Enzweiler, University of Applied Sciences Esslingen

SpatialDETR is a transformer-based approach for robust and scalable single-shot 3D object detection based on multi-view camera images. We use a DETR-like architecture with a 3D geometric positional encoding in combination with a spatially motivated, sensor-relative attention block. This leverages global context across sensor borders to detect objects present in the scene. Explicitly integrating the extrinsic calibration by computing the attention in a sensor-relative fashion allows to scale towards varying sensor sets or different sensor mounting positions.

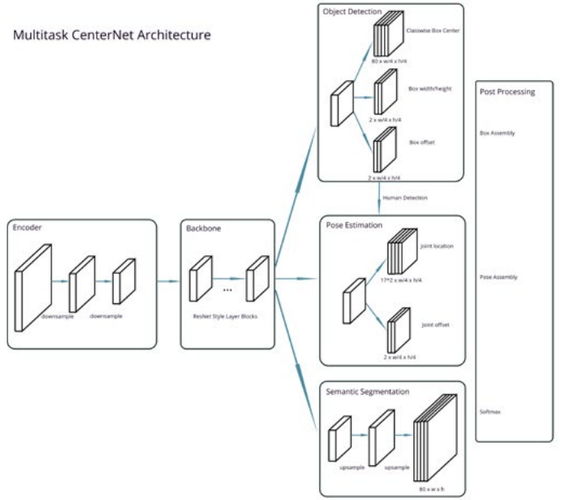


Visualization of the spatially-aware global cross-sensor attention. Fusing the 3D information from the individual images per object hypothesis is supported by a geometric positional encoding (left) that incorporates the global 3D view direction of each pixel. (© Mercedes-Benz | Motional AD Inc.)

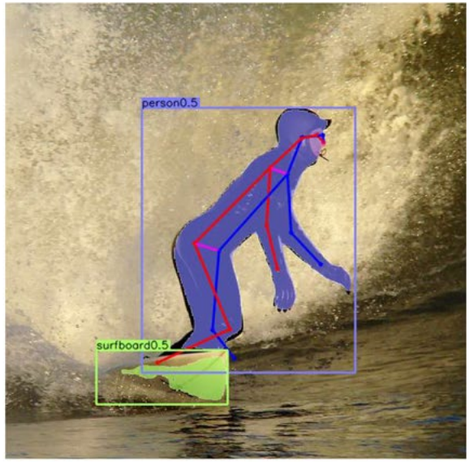
Knowledge Transfer for Multitask and Downstream Tasks

Falk Heuer, Sven Mantowsky, Syed Saqib Bukhari, Georg Schneider, ZF Group

Computer Vision algorithms in autonomous driving can only be safe when created in a robust, versatile and redundant manner. In order to implement them efficiently on limited hardware, multitask learning may use a shared backbone and safe valuable computational resources by jointly using layers of the neural network. Additionally, symbiotic effects can aid learning and help the network converge better on each individual task.
We train networks on the three tasks semantic segmentation, detection and human pose estimation and show that they can work together in an efficient manner. Additionally, we perform an analysis on various multitask setups and compare their joint performance to individual specialist networks.



Architecture of the proposed network. (© ZF Group)

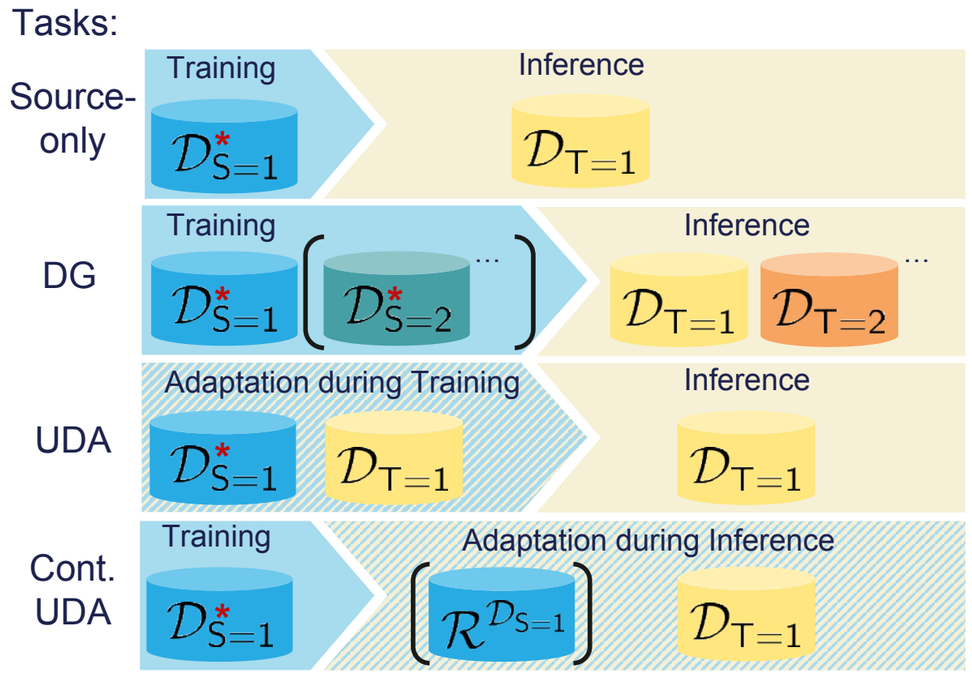


A multitask network predicting detections, segmentations and human pose estimation at the same time. (© ZF Group)

Domain Generalization and (Continuous) Unsupervised Domain Adaptation

Jan-Aike Termöhlen, Tim Fingscheidt, Technische Universität Braunschweig

We investigated methods for domain generalization (DG) and (continuous) unsupervised domain adaptation (UDA). We showed that the best performance on multiple target domains can be achieved with UDA methods. Additionally, some UDA methods are even suitable for domain generalization when adapted to the right “representative” target domains/data and perform better than state-of-the-art DG methods. Furthermore, we showed that with our continuous UDA it is possible to adapt online to the target domain during inference.

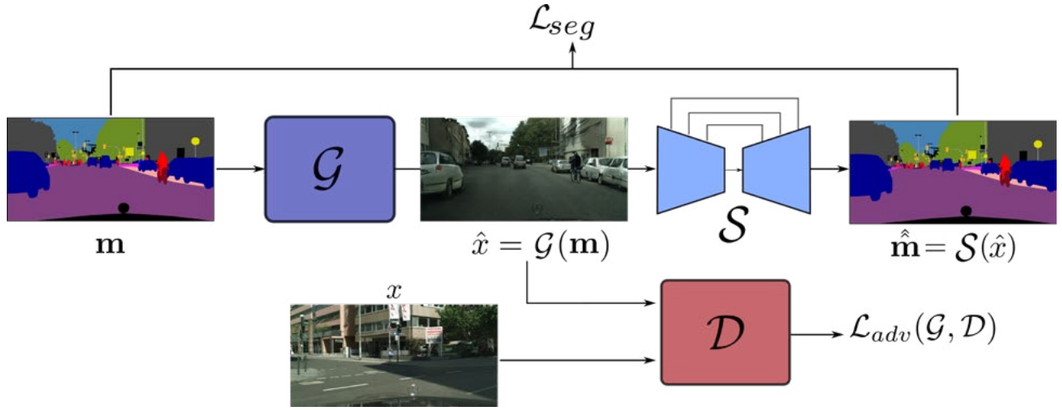


Simplified schematic visualization of different tasks regarding the domain gap. The red star indicates availability of labels. (© Technische Universität Braunschweig)

USIS: Unsupervised Semantic Image Synthesis

George Eskandar, Mohamed Abdelsamad, Karim Armanious, Diandian Guo, Bin Yang, University of Stuttgart
Shuai Zhang, University of Tübingen

Semantic Image Synthesis (SIS) is a subclass of I2I (I2) translation where a photorealistic image is synthesized from a segmentation mask. State-of-the-art methods depend on a massive amount of labeled data, while generic unpaired I2I frameworks underperform in comparison. In this work, we propose a new framework, Unsupervised Semantic Image Synthesis (USIS), as a first step toward closing the performance gap between paired and unpaired settings. USIS features a one-sided cycle-loss, a wavelet-based whole image discriminator, and a decoder on top of the discriminator to reconstruct the image and regularize the adversarial training. Moreover, we design a wavelet-based generator. USIS outperforms previous approaches in 3 challenging benchmarks.

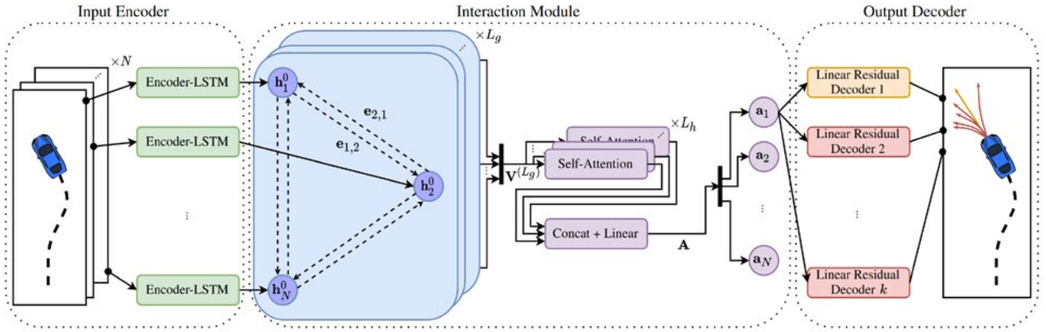


USIS can be trained on unpaired images and labels. It outperforms other unpaired frameworks by large margin in alignment, while delivering competitive image fidelity scores against supervised models.

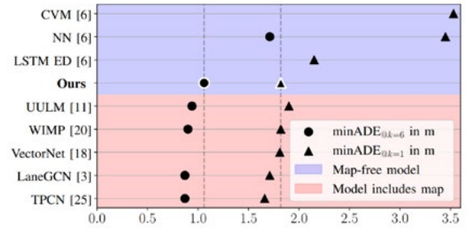
CRAT-Pred: Vehicle Trajectory Prediction with Crystal Graph Convolutional Neural Networks and Multi-Head Self-Attention

Julian Schmidt, Mercedes-Benz AG, Ulm University | **Julian Jordan, Franz Gritschneder**, Mercedes-Benz AG
Klaus Dietmayer, Ulm University

Map information is not always available. We propose CRAT-Pred, a multi-modal and non-rasterization-based trajectory prediction model, specifically designed to effectively model social interactions between vehicles, without relying on map information. CRAT-Pred applies a graph convolution using edge features, and combines it with multi-head self-attention. Compared to other map-free approaches, the model achieves state-of-the-art performance with a significantly lower number of model parameters. Additionally, the self-attention weights represent a measurable interaction score. The source code is publicly available.



Method	$k=1$			$k=6$		
	minADE	minFDE	MR	minADE	minFDE	MR
LSTM ED [6]	2.15	4.97	0.75	-	-	-
LSTM ED-soc. [6]	2.15	4.95	0.75	-	-	-
NN [6]	3.45	7.88	0.87	1.71	3.29	0.54
CVM [6]	3.53	7.89	0.83	-	-	-
Ours	**1.82**	**4.06**	**0.63**	**1.06**	**1.90**	**0.26**



Top: Model Architecture. Bottom left: Comparison to other map-free approaches. Bottom right: Comparison to map free and map aware models. (© Mercedes-Benz AG)

Semi- and Unsupervised Learning

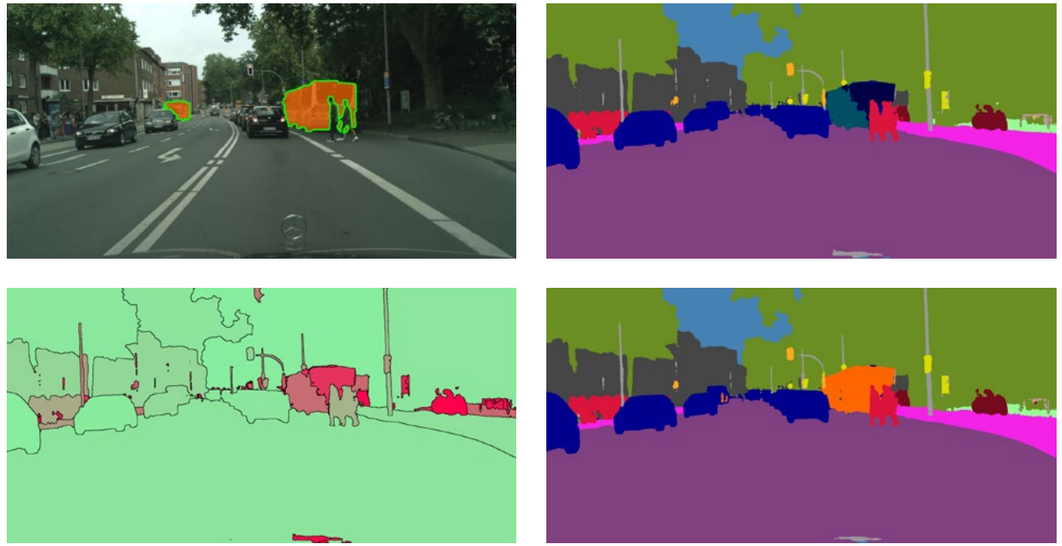
Re-training to adjust the pre-trained network to the domain of operation of the ego-car is conducted with only a few or without any labels of the target domain.

Towards Unsupervised Open World Semantic Segmentation 88
Semi-supervised domain adaptation with CycleGAN guided by downstream task awareness 90
Attention-Based Self-Supervised Monocular Depth Estimation 92
Cycle-Consistent World Models for Domain Independent Latent Imagination 94
3D-Aware Image Synthesis with Generative Radiance Fields 96
Augmentation-based Domain Generalization for Semantic Segmentation 98
Survey on Unsupervised Domain Adaptation for Semantic Segmentation for Visual Perception. 100
Self-Supervised Deep Representation Learning for Semantic Segmentation 102

Towards Unsupervised Open World Semantic Segmentation

Svenja Uhlemeyer, Matthias Rottmann, Hanno Gottschalk, University of Wuppertal

Open world semantic segmentation involves identifying pixels belonging to unknown objects and learning novel classes incrementally. It is desirable to perform such an incremental learning task in an unsupervised fashion. Our proposed method consists of four steps: anomaly segmentation, i.e., detecting and localizing unknown objects (1), which we cluster by visual similarity (2). Based on these clusters we create pseudo ground-truth for novel classes (3), by which we expand the segmentation model incrementally (4). In our experiments we demonstrate that, without access to ground-truth and even with few data, a deep neural network's class space can be extended by novel classes, achieving considerable segmentation accuracy.

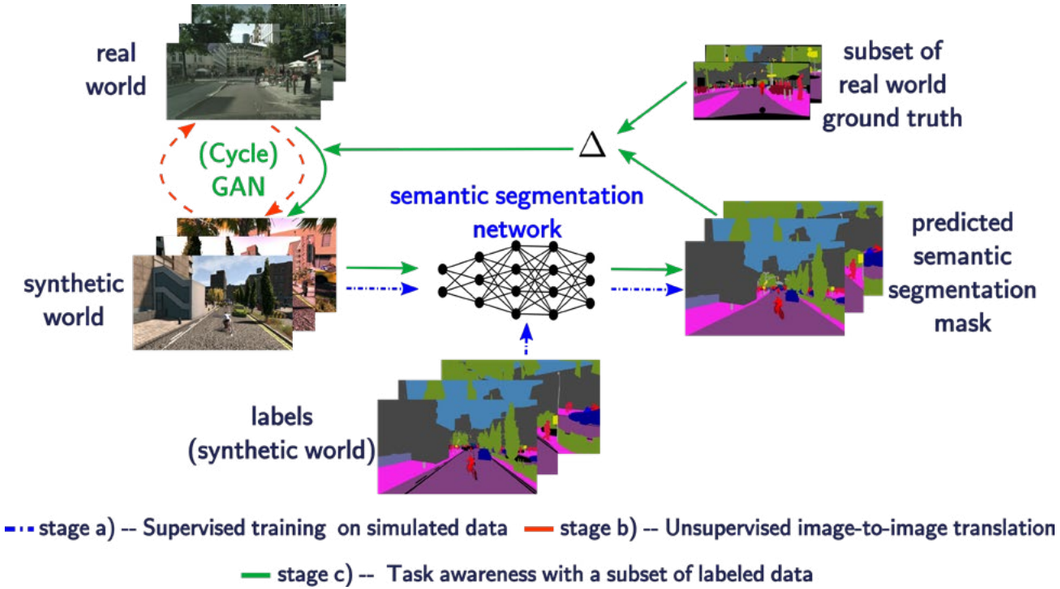


Top left: busses are annotated as unknown/novel class. Top right: The segmentation network assigns them to the known classes car, truck and train. Bottom left: the prediction quality estimation identifies one bus which may be badly predicted. Bottom right: this bus is recognized by the segmentation model after class-incremental learning.
(© Cityscapes dataset)

Semi-supervised domain adaptation with CycleGAN guided by downstream task awareness

Annika Mütze, Matthias Rottmann, Hanno Gottschalk, University of Wuppertal

Image-to-image (I2I) approaches can bridge domains on input level. Nevertheless, standard I2I approaches do not focus on the downstream task but rather on the visual inspection level. We therefore propose a “task aware” generative adversarial network in an I2I domain adaptation approach. Assisted by some labeled data, we guide the I2I translation to a more suitable input for a semantic segmentation network trained on synthetic data. This constitutes a modular semi-supervised domain adaptation based on CycleGAN. Our experiments involve evaluations on complex domain adaptation tasks and refined domain gap analyses using from-scratch-trained networks.

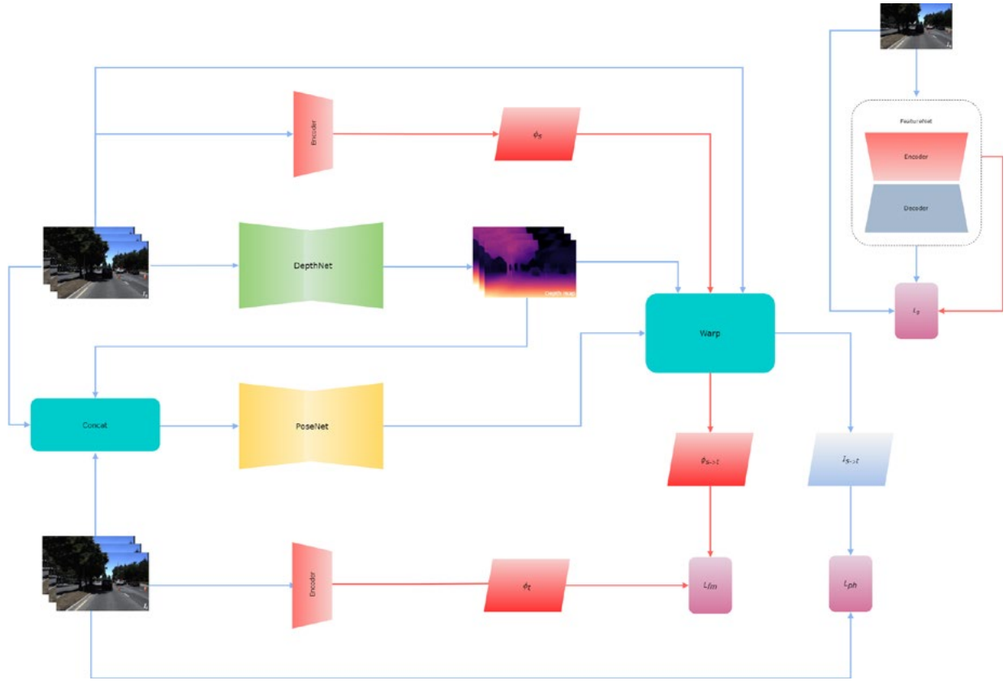


Concept of our method: Stage a) – Training of a downstream task model on the synthetic domain. Stage b) – Training a CycleGAN based on unpaired data to transfer real data into the synthetic domain. Stage c) – We freeze the downstream task network and tune the generator with the help of a few labeled data points by guiding it based on the loss of the downstream task network (© University of Wuppertal)

Attention-Based Self-Supervised Monocular Depth Estimation

Sagar Hanagodimath, Marius Bachhofer, ZF Group

Self-supervised monocular depth estimation approaches allow estimation of Euclidean distances of the objects from camera using only a sequence of images and no groundtruth. We studied the impact of using different attention mechanisms at various resolutions in the model and their influence on the overall model performance on real-world data. Furthermore, we also document the usefulness of incorporating additional high level feature maps and initial depth maps to improve the quality of the supervisory signals leading to better depth prediction.

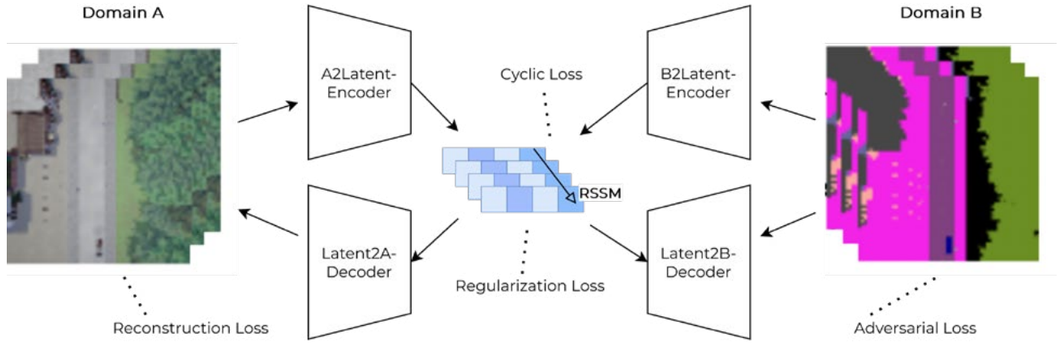


High-level overview of the method (Source and target images from DDAD. © Guizilini, Vitor, et al. „3d packing for self-supervised monocular depth estimation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.)

Cycle-Consistent World Models for Domain Independent Latent Imagination

Tim Joseph, FZI Forschungszentrum Informatik, Karlsruhe

For many tasks, such as autonomous driving, running an agent in the real world is prohibitively expensive and dangerous. For this reason, training the agent in a simulated environment before releasing it to the real world is necessary. However, many agents fail in the real world because they are not able to handle the domain gap between the real and the simulated environment. We present an approach that projects (sensor) data from two different domains into a common latent space. From the information conveyed in the latent representations, we can train a reinforcement learning agent that learns to drive, independently of the given sensor data.



The general architecture of our approach. Given two domains (e.g. RGB and Semseg) we can project to and from a common latent space. We can also forward predict in latent space to infer the future development of a scene.

3D-Aware Image Synthesis with Generative Radiance Fields

Katja Schwarz, Axel Sauer, Michael Niemeyer, Yiyi Liao, Andreas Geiger, University of Tübingen

While 2D generative neural networks enabled high-resolution image synthesis, they largely lack an understanding of the 3D world and the image formation process. To address this problem, **3D-aware generative adversarial networks** combine 3D generators, differentiable rendering and adversarial training to **synthesize novel images with explicit control over the camera pose** and, potentially, other scene properties like object shape and appearance.

3D-aware generative adversarial networks combine 3D generators, differentiable rendering and adversarial training to synthesize novel images with explicit control over the camera pose. Here, we show generated voxel grids. Note that the model is only trained with posed images and does not require multiview supervision.
(© University of Tübingen)

Augmentation-based Domain Generalization for Semantic Segmentation

Manuel Schwonberg, CARIAD SE | **Fadoua El Bouazati**, University of Wuppertal
Nico Schmidt, CARIAD SE | **Hanno Gottschalk**, University of Wuppertal

Unsupervised Domain Adaptation (UDA) and domain generalization (DG) are two research areas that aim to tackle the lack of generalization of Deep Neural Networks (DNNs) towards unseen domains. While UDA methods have access to unlabeled target images, domain generalization does not involve any target data and only learns generalized features from a source domain. We systematically study the in- and out-of-domain generalization capabilities of simple, rule-based image augmentations like blur, noise, color jitter and many more. On the challenging synthetic-to-real domain shift between Synthia and Cityscapes we reach 39.5% mIoU compared to 40.9% mIoU of the best previous work.

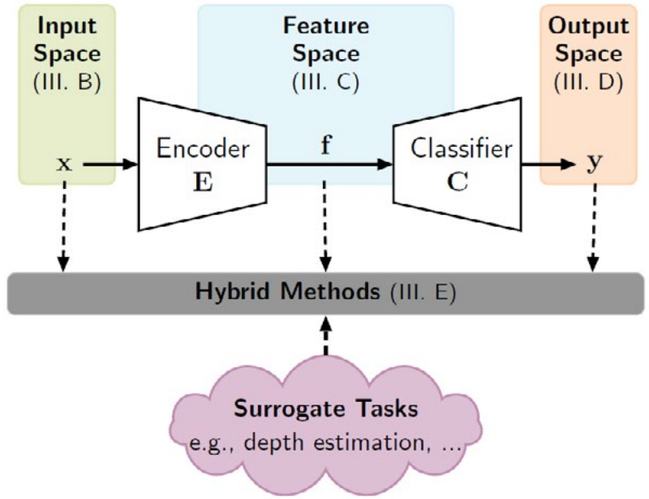
Method		Synthia
		to Cityscapes
Baseline (Ours)	ResNet-101	29.3 (29.6)
RandomCrop (Ours)		35.4 (35.7)
IBN [19]		34.2
SW [21]		31.6
DRPC [23]		37.6
GTR [27]		39.7
RobustNet [20]		37.2
FSDR [4]		40.8
AdvStyle [29]		37.6
WEDGE [28]		**40.9**
SAN&SAW [22]		**40.9**
RRCrop + ET (Ours)		37.8 (39.5)
Baseline (Ours)	DAFormer	39.6 (40.3)
RandomCrop		42.6
RRC,GB,CJitter		**44.2**

mIoU of different DG approaches in comparison with augmentation-based domain generalization (© CARIAD SE)

Survey on Unsupervised Domain Adaptation for Semantic Segmentation for Visual Perception

Manuel Schwonberg, CARIAD SE | **Joshua Niemeijer**, DLR
Jan-Aike Termöhlen, Technical University Braunschweig | **Jörg P. Schäfer**, DLR | **Nico Schmidt**, CARIAD SE
Hanno Gottschalk, University of Wuppertal | **Tim Fingscheidt**, Technical University Braunschweig

The bad generalization of Deep Neural Networks to new, unseen domains is a major problem on the way to a safe, large-scale application because manual annotation of new domains is costly. Methods are required to adapt DNNs to new domains without labeling effort: unsupervised domain adaptation (UDA). We present an overview of the current state of the art in this field of research. We categorize and explain the different approaches for UDA. We also present a quantitative comparison of the approaches.

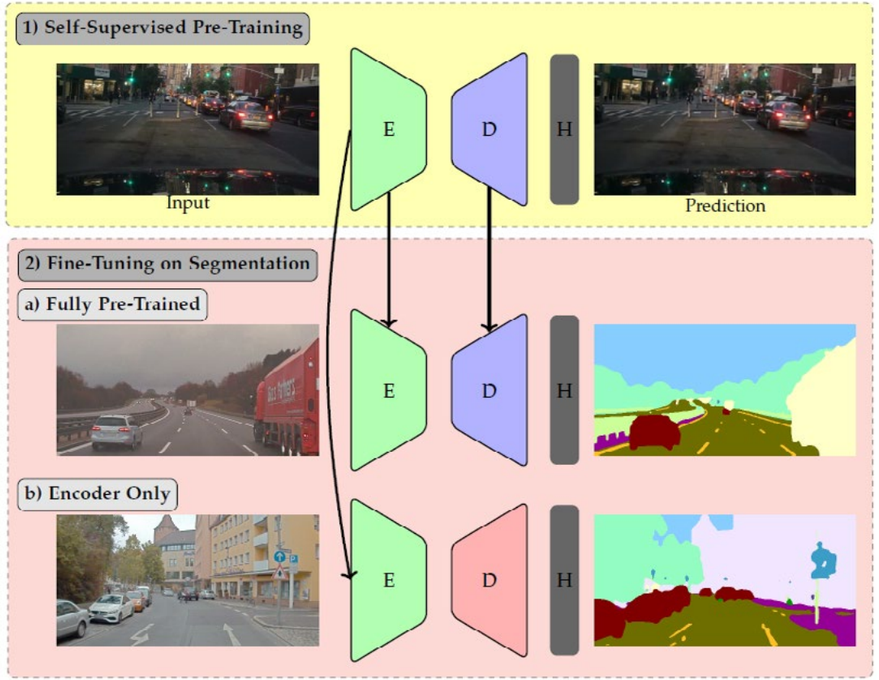


Taxoynomy of Unsupervised Domain Adaptation Approaches (© CARIAD SE, TU Braunschweig, DLR)

Self-Supervised Deep Representation Learning for Semantic Segmentation

Manuel Schwonberg, **Nico Schmidt**, CARIAD SE

Self-Supervised learning (SSL) methods aim to tackle the issue that Deep Neural Networks (DNNs) require large amounts of human-annotated data. The overall aim of this investigation is the evaluation of different self-supervised pretext tasks for automotive semantic segmentation and major influencing factors in comparison with the supervised ImageNet pre-training baseline. We evaluated a wide range of pretext tasks for semantic segmentation but none of them was capable to compete with supervised ImageNet pre-training. However, when including strong augmentations into the pre-training an significant improvement over random initializationwas observed.



Self-Supervised Pre-Training for Semantic Segmentation (© CARIAD SE)

Training Strategies

To obtain robust performance of neural networks, training is essential. There are many different strategies how to train neural networks. Contributions of the category Training Strategies investigate which strategies to be used for best results under given conditions.

PlanT: Explainable Planning Transformers via Object-Level Representations 106
Automated Detection of Label Errors in Semantic Segmentation Datasets 108
Severity of Catastrophic Forgetting in Object Detection for Autonomous Driving 110
MGiaD: Multigrid in all dimension. Efficiency and Robustness by Coarsening in Resolution and Channel Dimensions. 112
Improving Replay-Based Continual Semantic Segmentation with Smart Data Selection 114
Causes of Catastrophic Forgetting in Class-Incremental Semantic Segmentation. 116

PlanT: Explainable Planning Transformers via Object-Level Representations

Katrin Renz, Kashyap Chitta, Otniel-Bogdan Mercea, A. Sophia Koepke, Zeynep Akata, Andreas Geiger,
University of Tübingen

We propose PlanT, a novel approach for planning in the context of self-driving that uses a standard transformer architecture. PlanT is based on imitation learning with a compact object-level input representation. On the Longest6 benchmark for CARLA, PlanT outperforms all prior methods (matching the driving score of the expert) while being 5.3× faster than equivalent pixel-based planning baselines during inference. Furthermore, we propose an evaluation protocol to quantify the ability of planners to identify relevant objects, providing insights regarding their decision-making. Our results indicate that PlanT can focus on the most relevant object in the scene, even when this object is geometrically distant.

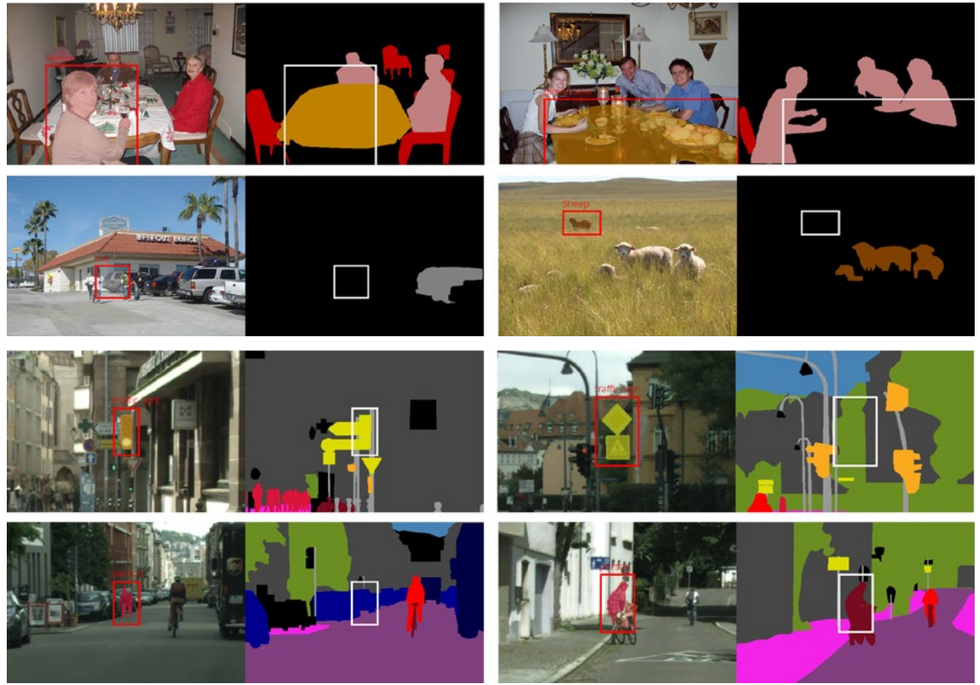
PlanT can drive safely in urban environments and in addition we can visualize on which vehicles the decision was based on. Vehicles with the highest relevance score are marked with a red bounding rectangle. We show examples for successful matching of relevance score and intuition (green frames) and failures (red frames). (© University of Tübingen)

Automated Detection of Label Errors in Semantic Segmentation Datasets

Matthias Rottmann, Marco Reese, University of Wuppertal

We present a method and a benchmark for the detection of label errors in semantic segmentation datasets [1]. For the detection, we utilize deep learning and uncertainty quantification. If a predicted segment is a false positive w.r.t. the ground truth while our uncertainty estimate signals high confidence, then we review this finding. We studied the efficiency of our method on our benchmark in detail and found plenty of label errors in state-of-the-art computer vision datasets.

[1] Rottmann, Matthias, and Marco Reese. „Automated Detection of Label Errors in Semantic Segmentation Datasets via Deep Learning and Uncertainty Quantification.“ Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). 2023.



Label Errors found in the datasets Pascal VOC (top) and Cityscapes (bottom). In each subfigure, the left panel shows our detection and the right hand panel the ground truth

Severity of Catastrophic Forgetting in Object Detection for Autonomous Driving

Christian Witte, Syed Saqib Bukhari, Georg Schneider, ZF Group

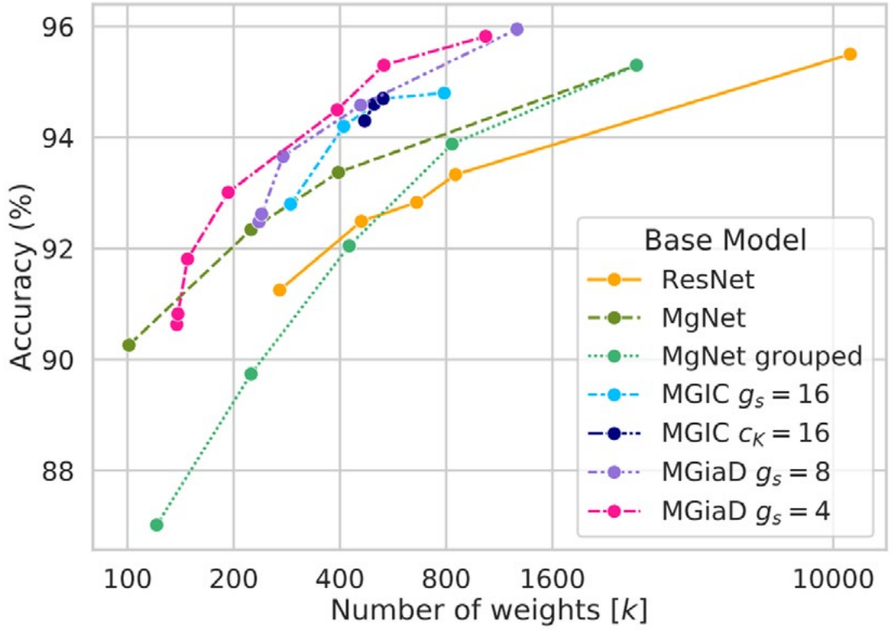
Our work aims to illustrate the severity of catastrophic forgetting for object detection for class- and domain-incremental learning. We propose four hypotheses, as we investigate the impact of the ordering of sequential increments and the underlying data distribution of AD datasets. Further, the influence of different object detection architectures is examined. The results of our empirical study highlight the major effects of forgetting for class-incremental learning. Moreover, we show that domain-incremental learning suffers less from forgetting but is highly dependent on the design of the experiments and choice of architecture.

Class-incremental (a) and Domain-incremental (b) Learning for Object Detection and for the BDD100K Dataset (© ZF Group)

MGiaD: Multigrid in all dimension. Efficiency and Robustness by Coarsening in Resolution and Channel Dimensions

Antonia van Betteray, Matthias Rottmann, Karsten Kahl, University of Wuppertal

Current deep neural networks (DNNs) for image classification are made up of 1000 million learnable parameters. Despite their high classification accuracy these networks are heavily overparameterized. Active research in recent years in terms of using multigrid inspired ideas in DNNs have shown that on one hand a significant number of weights can be saved by appropriate weight sharing and on the other that a hierarchical structure in the channel dimension can improve the weight complexity to linear. Utilizing these findings, we introduce an architecture that establishes multigrid structures in all relevant dimensions, contributing a drastically improved accuracy-parameter trade-off.

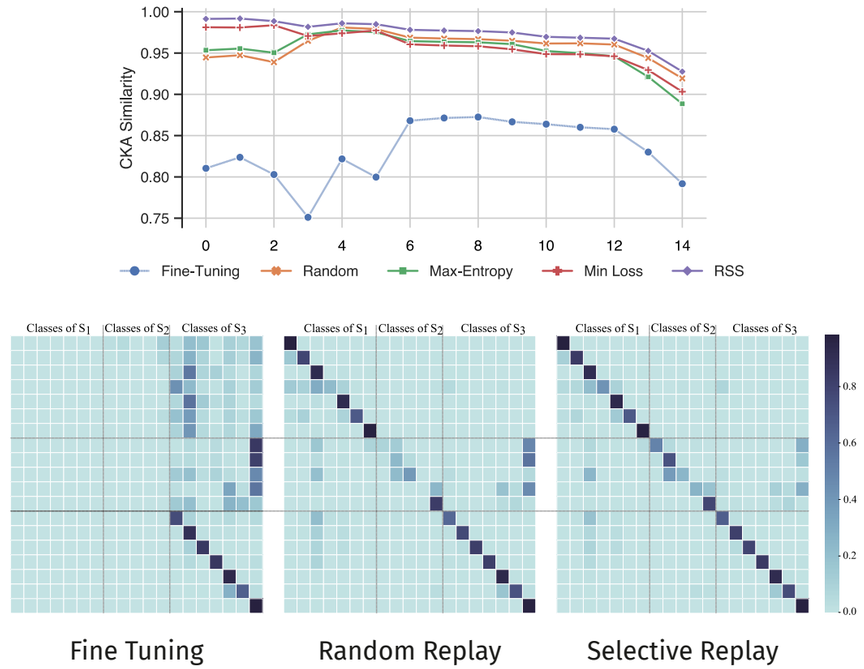


Accuracy-parameter trade-off of MGiaD models trained on CIFAR-10. The group size is fixed and multigrid parameters vary. Corresponding ResNet and MgNet as well as MGIC as benchmarks

Improving Replay-Based Continual Semantic Segmentation with Smart Data Selection

Tobias Kalb, Porsche Engineering Group GmbH

Replay has proven to be effective in reducing forgetting for Continual Semantic Segmentation. The most common sample strategy for Replay is random selection, which can result in unstable results. Therefore, we investigate the influences of various replay strategies for semantic segmentation and evaluate them in class- and domain-incremental settings. Our results show that effective sampling methods help to decrease the representation shift in early layers and the task recency bias, which are both a major cause of forgetting in domain-incremental learning.

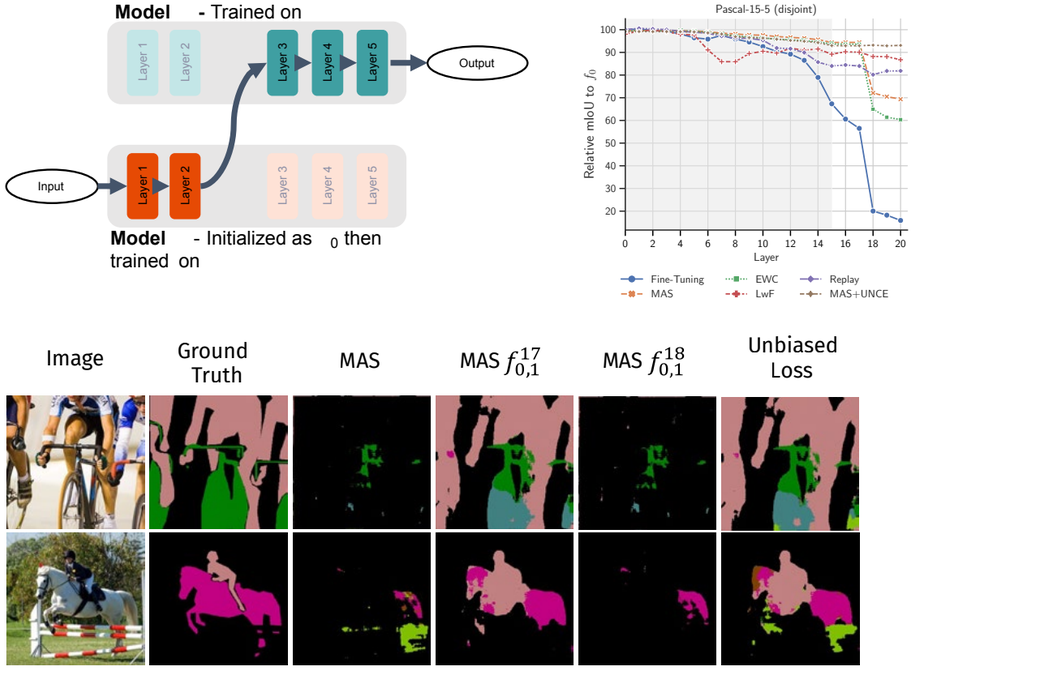


Top: Representational similarity (CKA) between activations of all layers (horizontal axis) before and after incremental training.
Bottom: Confusion Matrix after Incremental Training. Note the miss-classification for non-reoccurring classes of S2 and the improved accuracy after employing selective replay

Causes of Catastrophic Forgetting in Class-Incremental Semantic Segmentation

Tobias Kalb, Porsche Engineering Group GmbH

We study the causes of catastrophic forgetting in Class-Incremental Semantic Segmentation, answering how it manifests itself in the hidden representations of the network and how the background class both causes severe forgetting and decreases activation drift. Using representational similarity techniques, we demonstrate that forgetting manifest itself in deeper layers of the networks by assigning previous discriminating features for the previous classes to the background class or visually related classes. However, re-appearing classes mitigate activation drift in the encoder even when they are labelled as background.



By using layer stitching (upper left) at specific layers, we measure the relative mIoU of the stitched network to measure the activation drift. Our results (upper right) show that deeper layers are the main causes of forgetting. The predictions of the stitched networks (bottom) show that the old classes are assigned to the background class in later layers.

Embedded Systems

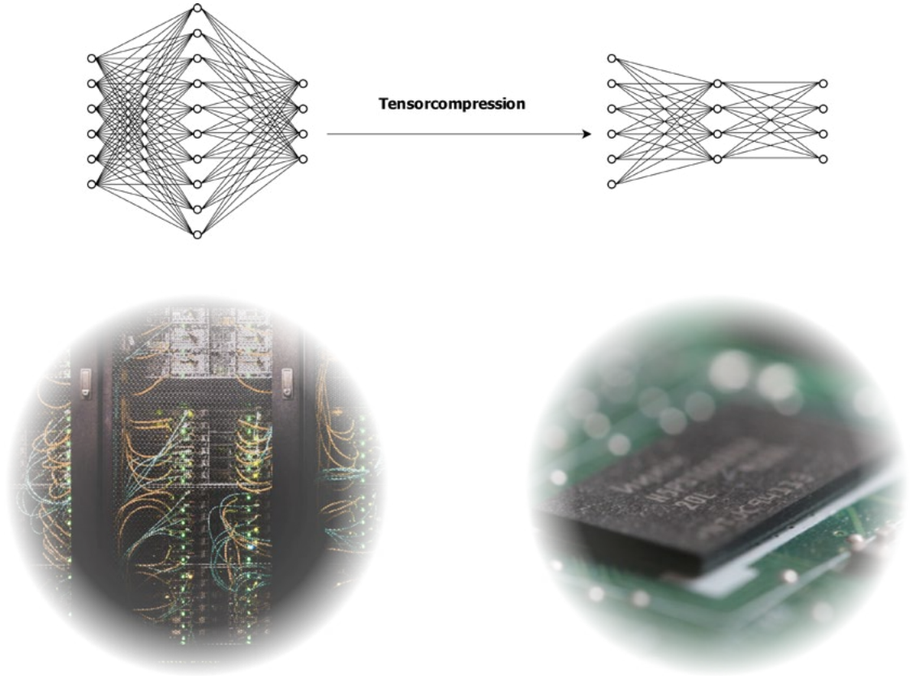
AI systems are trained and executed in the lab on powerful hardware. However, when used in a vehicle, they must function with very limited resources, especially computing speed and memory.

Analyzing and optimizing AI for embedded applications 120
Interpretable Pruning 122
Accelerating and Pruning CNNs for Semantic Segmentation on FPGA 124

Analyzing and optimizing AI for embedded applications

Domenik Helms, Adrian Osterwind, Arunachalam Thirunavukkarasu,
Deutsches Zentrum für Luft- und Raumfahrt e.V.

When deploying Artificial Neural Networks on embedded AI, one faces resource restrictions. Because of this it is useful knowing the resource requirements of a neural network before training and deploying to optimize beforehand. This is why we worked on a prediction methodology for the final execution time of a neural network.If one on the other hand already has a functioning network which does not meet these requirements, optimizations have to be performed while keeping training at a minimum. We thus worked on tensor compression, which aims to reduce the size of a matrix while keeping the result of the operation the same. Since the compression has many parameters an automated parameter search based on a network architecture search heuristic was also implemented.

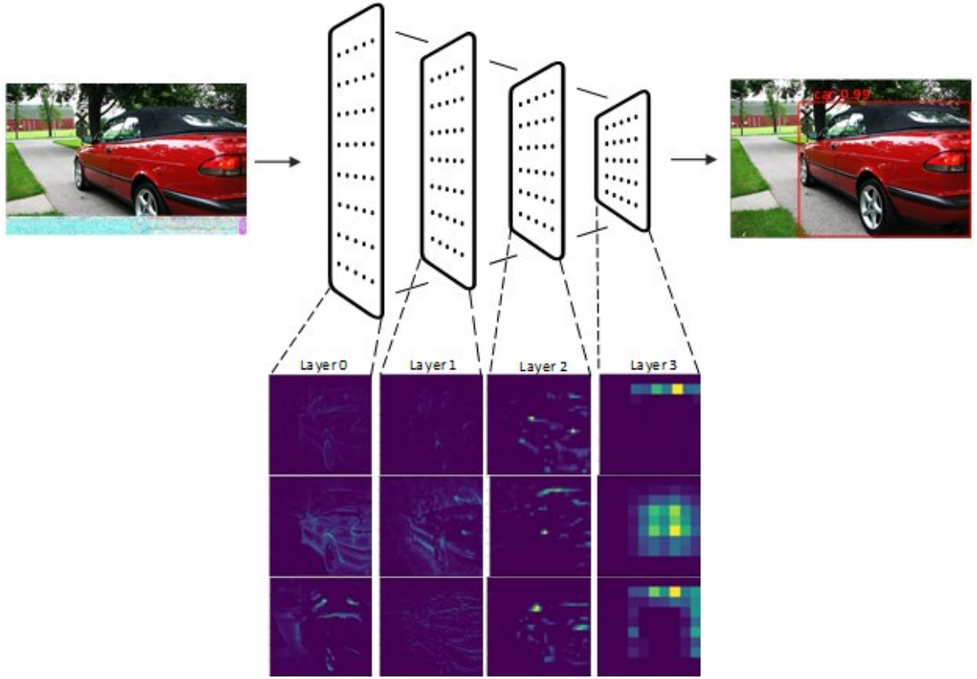


Tensor compression to allow for the delta between a training server and a target device (© Unsplash Inc., Pixabay GmbH)

Interpretable Pruning

Sven Mantowsky, Syed Saqib Bukhari, Georg Schneider, ZF Group

Pruning is a technique to remove less important neurons/filters from a model, making it more efficient while preserving or improving performance. This process usually is non-transparent for the user and is hardly interpretable. ZF has faced this problem and developed a method that combines pruning and interpretability, called Interpretable Pruning. Using heatmaps generated by Deep Taylor Decomposition, the user can understandably evaluate which filters contribute the most to the predictions. The method generates a ranking based on these heatmaps and the user can determine the number of filters to be removed based on the ranking. We have tested this method with both classification and object detection. With the object detector SSD and the PASCAL VOC dataset we could achieve a compression rate of 40%, with the classifier VGG16 and the CIFAR100 dataset a rate of over 70%.

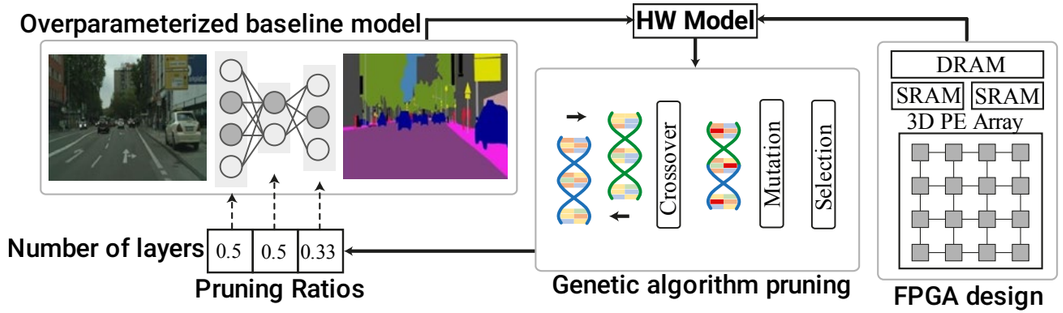


Representation of the heatmaps of different layers used by the Interpretable Pruning method (note the decreasing resolution with increasing network depth). (© ZF Group)

Accelerating and Pruning CNNs for Semantic Segmentation on FPGA

Manoj Vemparala, Alexander Frickenstein, Lukas Frickenstein, BMW AG | **Nael Fasfous, Pierpaolo Mori, Saptarshi Mitra**, Technical University of Munich

We establish an end-end deployment pipeline for semantic segmentation using channel pruning and HW model (See Figure). We formulate the channel pruning as search problem using genetic algorithm, where redundant filters are pruned based on layer-wise compression ratios and a magnitude-based heuristic.
Proxy metrics, such as operation count (OPs), does not always guarantee tangible improvements on measured hardware estimates. In our results the Hardware Aware pruning outperforms OPs based pruning both in latency and compute complexity at equal mIoU.



Depiction of the end-to-end CNN deployment pipeline on an embedded platform using genetic search based channel pruning.

Real World Robustness

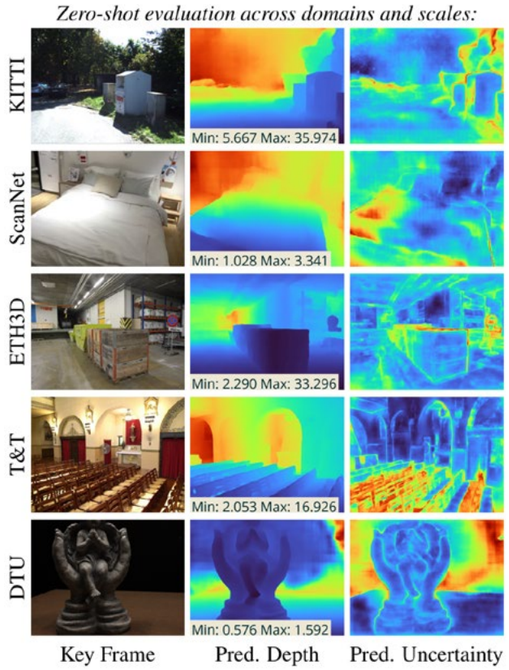
AI systems are trained in the lab on previously recorded or artificially generated data. In actual use in the vehicle, however, unexpected and unknown situations may arise.

A Benchmark and a Baseline for Robust Multi-view Depth Estimation 128
Unsupervised Detection of Abnormal Driving Behavior. 130
Impact of Data Anonymization of Semantic Segmentation 132

A Benchmark and a Baseline for Robust Multi-view Depth Estimation

Philipp Schröppel, Artemij Amiranashvili, Thomas Brox, University of Freiburg
Jan Bechtold, Robert Bosch GmbH

We introduce the Robust Multi-View Depth Benchmark that is built upon a set of public datasets and allows evaluation in depth-from-video and multi-view stereo settings on data from different domains. We evaluated recent approaches and found imbalanced performances across domains. Further, we considered a third setting, where the objective is to estimate the corresponding depth maps with their correct scale. We could show that recent approaches do not generalize across datasets in this setting. To resolve this, we present the Robust MVD Baseline model for multi-view depth estimation, which is built upon existing components but employs a novel scale augmentation procedure.

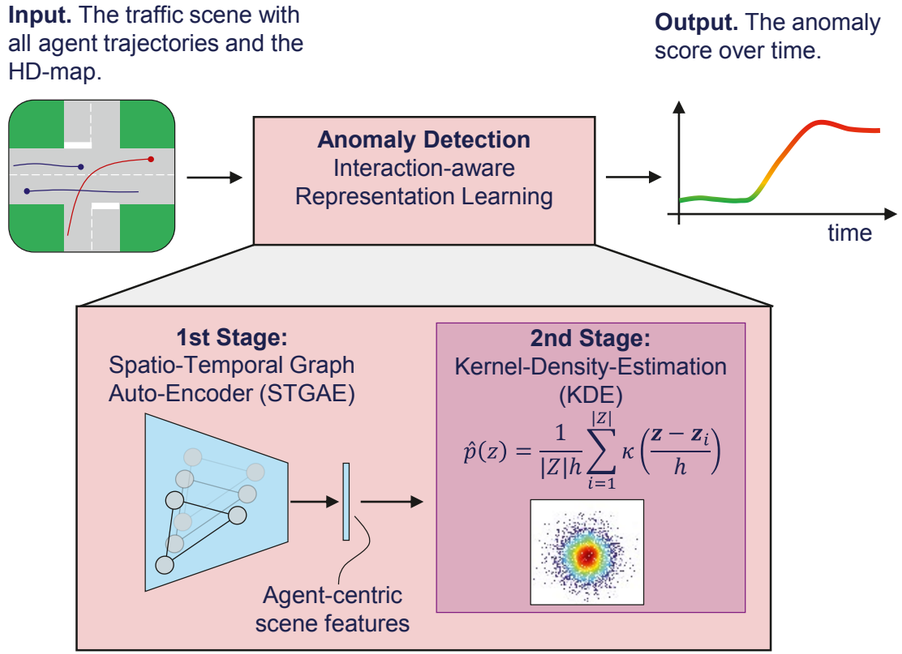


The Robust Multi-View Depth Benchmark evaluates robust multi-view depth estimation on arbitrary real-world data. As a proxy it defines test sets based on multiple diverse existing. This simulates an open-world scenario where it is always possible to encounter scenarios not covered by the training data. (© University of Freiburg)

Unsupervised Detection of Abnormal Driving Behavior

Julian Wiederer, Julian Schmidt, Arij Bouazizi, Ulrich Kreßel, Mercedes-Benz AG
Vasileios Belagiannis, Friedrich-Alexander-University Erlangen-Nürnberg

Human intuition allows to detect abnormal driving scenarios in situations they never experienced before. Like humans detect abnormal situations and take counter-measures to prevent collisions, self-driving cars need anomaly detection mechanisms. We propose the R-U-MAAD benchmark for unsupervised anomaly detection in multi-agent trajectories. To this end we combine a replay of real-world trajectories and scene-dependent abnormal driving in the simulation. We learn a probability distribution of the normal driving from the training sequences without labels, and afterwards detect anomalies in low-density regions.

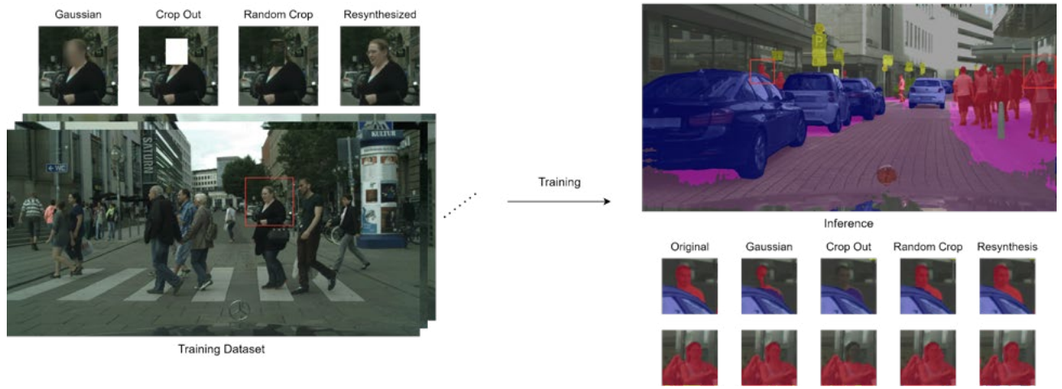


Two stage detection of driving anomalies. **1st Stage.** Extract agent-centric feature vectors with a spatio-temporal graph encoder (STGAE) from the given traffic scene. **2nd Stage.** Compute the anomaly score as the similarity between the feature vector the training data using a Kernel Density Estimation (KDE). (© Mercedes-Benz AG)

Impact of Data Anonymization of Semantic Segmentation

Jingxing Zhou, Porsche Engineering Group GmbH | **Jürgen Beyerer**, Fraunhofer IOSB and KIT

For the development of machine learning-based driver assistance systems and highly automated driving functions, training data play a significant role in ensuring machine learning algorithms generalize well on real driving scenarios. However, before camera images save on a server, license plates and faces of individuals should be anonymized first due to data privacy regulations. Nevertheless, the impact of using anonymized data on the performance of machine learning algorithms remains unclear. Our work aims to evaluate the impact of anonymization on the task of semantic segmentation using diverse neural network architectures, a range of input image resolutions, and different anonymization patterns. We observe statistically significant effects of anonymizing image data on model performance and investigate methods for mitigating segmentation precision loss.



The pipeline of our evaluation setup. Neural networks with different backbones and decoders are trained with identical training setup using diverse anonymization patterns. The segmentation images from ResNet 18 based FCN network are shown as examples. Images are anonymized using image resynthesis.

Sub-Project and Work Package Leads

SP1	**Christian Witt** | **Valeo Schalter und Sensoren GmbH**
WP1.1	Roshan Muthaiya | **CMORE Automotive GmbH**
WP1.2	Tobias Wagner, Christian Witt **Valeo Schalter und Sensoren GmbH**
WP1.3	Thies de Graaff | **DLR**
WP1.4	Jörg P. Schäfer | **DLR**

SP2	**Jens Mehnert** | **Robert Bosch GmbH**
WP2.1	Sebastian Wirkert | **BMW Group**
WP2.2	Martin Simon | **Valeo Schalter und Sensoren GmbH**
WP2.3	Jens Mehnert | **Robert Bosch GmbH**
WP2.4	Manuel Schwonberg | **CARIAD SE**
WP2.5	Florian Piewak | **Mercedes-Benz AG**

SP3	**Marius Bachhofer, Saqib Bukhari (Co-Lead)** **ZF Friedrichshafen AG** **Hanno Gottschalk** | **University of Wuppertal**
WP3.1	Tobias Kalb | **Porsche AG**
WP3.2	Antonia van Betteray | **University of Wuppertal**
WP3.3	Sebastian Wirkert, Stefan Matthes | **BMW Group**
WP3.4	Viviane Benzin | **Mercedes-Benz AG**

SP4	**Domenik Helms** | **DLR**
WP4.1	Julian Wiederer | **Mercedes-Benz AG**
WP4.2	Domenik Helms | **DLR**

Sub-Project and Work Package Leads

SP5	**Amin Hosseini | Mercedes-Benz AG**	**SP6**	**Amin Hosseini | Mercedes-Benz AG**
WP5.1	Amin Hosseini | **Mercedes-Benz AG**	WP6.1	Amin Hosseini | **Mercedes-Benz AG**
WP5.2	Sebastian Wirkert | **BMW Group**	WP6.2	Amin Hosseini | **Mercedes-Benz AG**
WP5.3	Franz Andert | **DLR**	WP6.3	Amin Hosseini | **Mercedes-Benz AG**

Table of contents

Welcome **2**

Greeting **4**

Collaboration in Artificial Intelligence **6**

KI Delta Learning **8**

Key Facts **10**

Data **14**

Transfer Learning **16**

Didactics **18**

Automotive Suitability **20**

Environment **22**

Improving robustness against common corruptions with frequency biased models **24**

Introducing Intermediate Domains for Effective Self-Training during Test-time **26**

Robustness Against Noisy Labels Through Uncertainty Estimation for LiDAR-based Semantic Segmentation **28**

An Unsupervised Domain Adaptive Approach for Multimodal 2D Object Detection in Adverse Weather Conditions **30**

A Low-Complexity Approach for Domain Adaptation **32**

Continual Learning for Model-Based Reinforcement Learning **34**

Motion Capture-based Virtual Reality Co-Simulation **36**

Domain Shift Quantification using Activations **38**

SceneNeRF: 3D Reconstruction of Real-World Scenes **40**

Environmental adaptation and self-attention in the context of unsupervised domain adaptation **42**

Detection of critical weather situations in scenario-based traffic simulations using optimization techniques **44**

Sensors **46**

3D Detection and Tracking From LiDAR Point Clouds As a Pre-Processing Step for Active Learning **48**

Processing of vehicle sensor data **50**

Real Data Acquisition with Ground Truth **52**

Auxiliary Task-Guided CycleGAN for Black-Box Model Domain Adaptation **54**

Bridging Domain Gaps in Lidar Perception **56**

Lidar Upsampling with Sliced Wasserstein Distance **58**

TransFuser: Imitation with Transformer-Based Sensor Fusion **60**

HALS: A Height-aware Lidar Super-Resolution Approach for Autonomous Driving **62**

Active Learning **64**

Active Learning On Dynamic Scenes Using Multi-View Consistency **66**

Active Learning based on a Taxonomy for Scene Description **68**

Active learning for semantic segmentation in realistic driving scenarios **70**

Consistency-based Active Learning for Semantic Segmentation **72**

Knowledge Transfer **74**

SpatialDETR: 3D Object Detection from Multi-View Camera Images with Global Cross-Sensor Attention **76**

Knowledge Transfer for Multitask and Downstream Tasks **78**

Domain Generalization and (Continuous) Unsupervised Domain Adaptation **80**

USIS: Unsupervised Semantic Image Synthesis **82**

CRAT-Pred: Vehicle Trajectory Prediction with Crystal Graph Convolutional Neural Networks and Multi-Head Self-Attention **84**

Semi- and Unsupervised Learning **86**

Towards Unsupervised Open World Semantic Segmentation **88**

Semi-supervised domain adaptation with CycleGAN guided by downstream task awareness **90**

Attention-Based Self-Supervised Monocular Depth Estimation **92**

Cycle-Consistent World Models for Domain Independent Latent Imagination **94**

3D-Aware Image Synthesis with Generative Radiance Fields **96**

Augmentation-based Domain Generalization for Semantic Segmentation **98**

Survey on Unsupervised Domain Adaptation for Semantic Segmentation for Visual Perception **100**

Self-Supervised Deep Representation Learning for Semantic Segmentation **102**

Training Strategies **104**

PlanT: Explainable Planning Transformers via Object-Level Representations **106**

Automated Detection of Label Errors in Semantic Segmentation Datasets **108**

Severity of Catastrophic Forgetting in Object Detection for Autonomous Driving **110**

MGiaD: Multigrid in all dimension. Efficiency and Robustness by Coarsening in Resolution and Channel Dimensions **112**

Improving Replay-Based Continual Semantic Segmentation with Smart Data Selection **114**

Causes of Catastrophic Forgetting in Class-Incremental Semantic Segmentation **116**

Embedded Systems **118**

Analyzing and optimizing AI for embedded applications **120**

Interpretable Pruning **122**

Accelerating and Pruning CNNs for Semantic Segmentation on FPGA **124**

Real World Robustness **126**

A Benchmark and a Baseline for Robust Multi-view Depth Estimation **128**

Unsupervised Detection of Abnormal Driving Behavior **130**

Impact of Data Anonymization of Semantic Segmentation **132**

Sub-Project and Work Package Leads **134**

Sub-Project and Work Package Leads **136**

Table of contents **138**

Contact & Further Information **140**

Contact & Further Information

Project Coordination

Dr.-Ing. Amin Hosseini
Mercedes-Benz AG, Research & Development
Function und Software Urban Autonomous Driving
HPC G140
71063 Sindelfingen, Germany
amin.hosseini@mercedes-benz.com

Deputy Project Coordinator

Marius Bachhofer
ZF Friedrichshafen AG
System House Autonomous Mobility Systems
88038 Friedrichshafen, Germany
marius.bachhofer@zf.com

Project Management

European Center for Information and
Communication Technologies – EICT GmbH
EUREF Campus Haus 13
Torgauer Straße 12-15
10829 Berlin, Germany

Poster Download

www.ki-deltalearning.de/finalevent

Project Consortium

Mercedes-Benz · BMW GROUP · MINI · BOSCH · CARIAD · Porsche Engineering driving technologies · Valeo · ZF · Luxoft · InnoSenT · DLR · FZI · Bergische Universität Wuppertal · Hochschule Reutlingen Reutlingen University · Technische Universität München · TUM · Universität Freiburg · Universität Stuttgart · Eberhard Karls Universität Tübingen

External Technology Partners

Technische Universität Braunschweig · DXC Technology · universität ulm · eict

Note on the legal form of the cooperation
The cooperation between the partners within the project has no independent legal personality. In fact a scientific exchange is conducted between the research centers, organizations and universities listed as cooperation partners. A legal or similar relationship under company law, an association or similar is not established by the scientific cooperation. No cooperation partner is entitled to represent individual other cooperation partners or all cooperation partners together towards third parties.