AI as Key Technology for Pattern Recognition and Situational Awareness

How does the Vehicle get the Knowledge it needs?

Franz Andert, Joshua Niemeijer, Jörg P. Schäfer, Gurucharan Srinivas
Institute of Transportation Systems
Berlin/Braunschweig

Task Context: Pattern Recognition in Sensor Data for Automatic Driving

From Low-Level Sensor Data to understandable High-Level Information

Objects

- road surface, markers, borders, traffic signs
- vehicles and their type
- pedestrians and other road users
- obstacles, dangers
- …

Properties of objects

- location (w.r.t. sensor, vehicle, and world)
- meaning (e.g. traffic sign type and light status)
- context-sensitive meaning (e.g. traffic rules)
- behavior (e.g. movement prediction)
- …

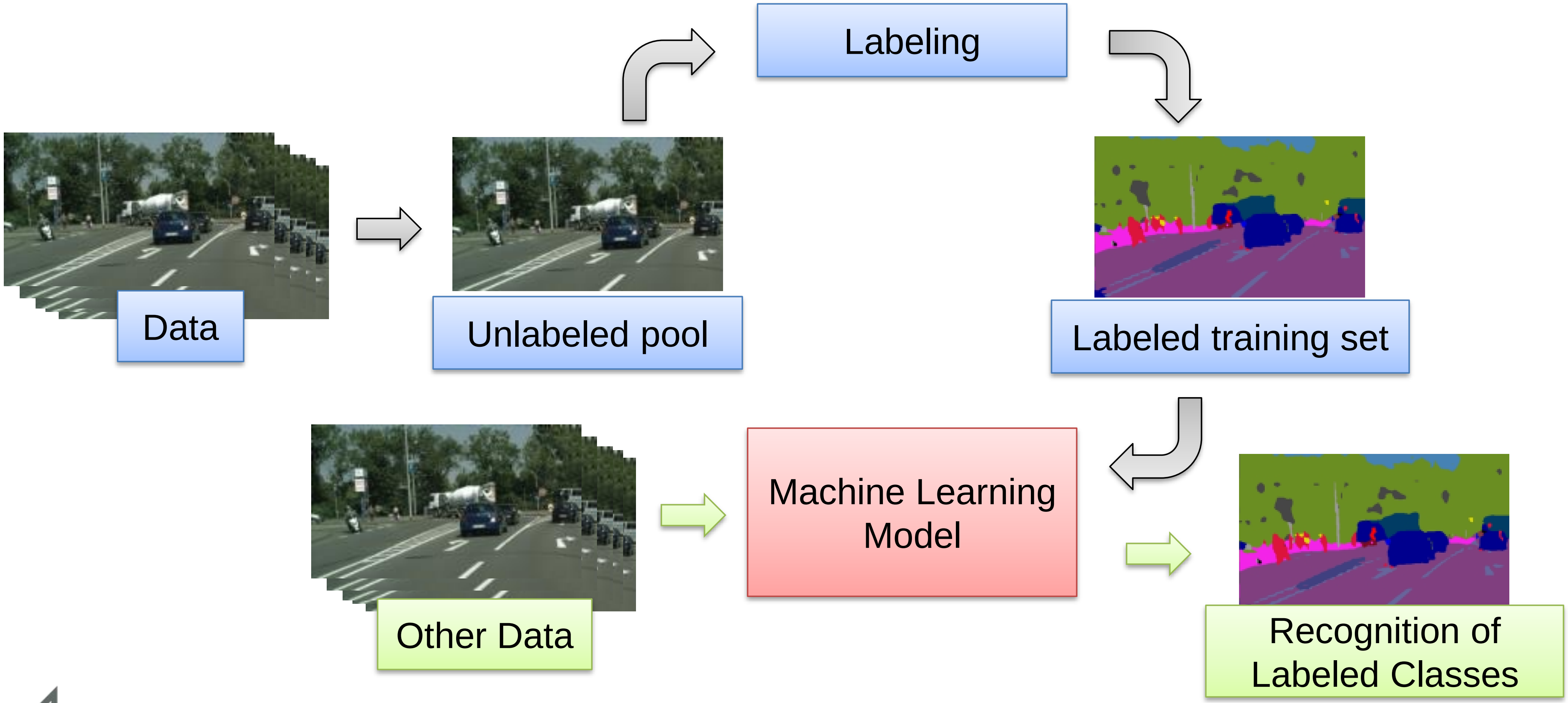
Inference to own driving

- vehicle ego-localization (w.r.t. world and environment)
- situational awareness (e.g. context-sensitive driving behavior)
- …

Patterns in Image Data: often too Complex for Manual Parameterization

AI / Deep Learning might be THE answer…

Supervised Learning seems to be simple:



Images from CityScapes Dataset –
Cordts et al., CVPR 2016

Patterns in Image Data: often too Complex for Manual Parameterization

AI / Deep Learning might be THE answer…

State-of-the-art example, trained with Cityscapes data

Use on Further Cityscapes Set [1]

Use on BDD Set [2]

[1] Cordts et al., The Cityscapes Dataset for Semantic Urban Scene Understanding, CVPR 2016, www.cityscapes-dataset.com
[2] Yu et al., BDD100K: A Diverse Driving Dataset for Heterogeneous Multitask Learning, CVPR 2020, www.bdd100k.com

Patterns in Image Data: often too Complex for Manual Parameterization

AI / Deep Learning might be THE answer…

Not optimal for (e.g.) camera images, not optimal for driving scenarios:

- ML model will be complex (Gigabytes+)
- Required data amount is HUGE (Terabytes+)
- Manual Labeling is VERY expensive (clickwork, crowdsourcing)

… with the result:

- even training can be expensive (GPU super-computing etc.)
- training data might be not complete
- the ML model is neither optimal nor complete
- the ML model may not contain context (“invisible knowledge”)

… so we need:

- reduce labeling cost (find most significant data to train with)
- data automation (generate synthetic data, with ground truth)
- labeling automation (semi-/unsupervised learning)
- extend model with various new data (domain adaptation)
- models to include context (knowledge integration)

Measure scenario coverage, prefer to add data from unknown and critical cases

Suitable rendering available, ~99% of the training data will be synthetic

Automation with human-in-the loop (active learning)

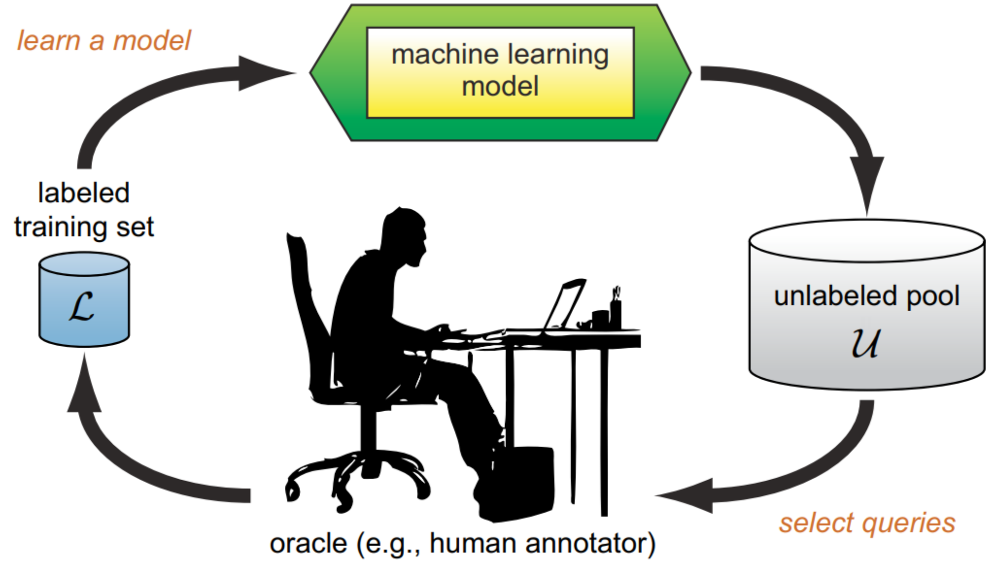
Automatic model extension benefits when domain shift is small

Use ontology of mutual object relations and in-visible knowledge

Reduce Labeling Cost with Active Learning

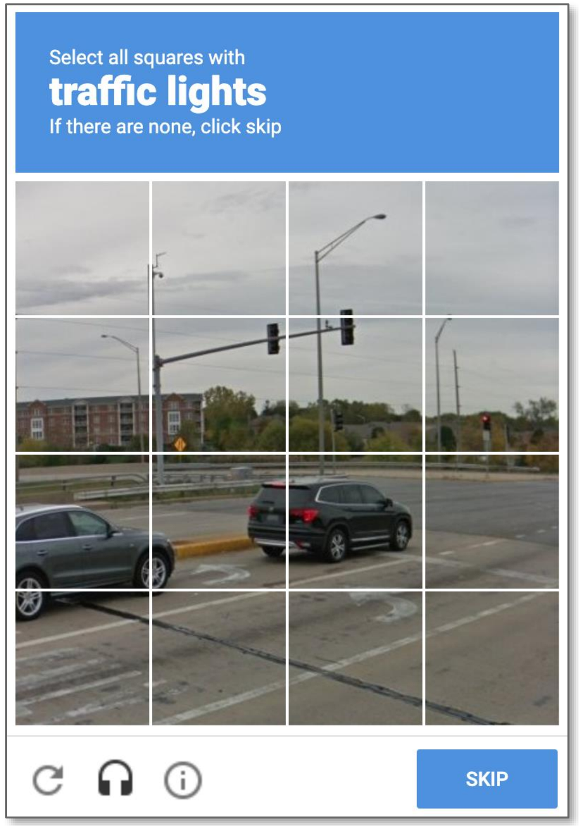
Strategy to find good data for training

General idea: algorithm queries user (oracle) [3]

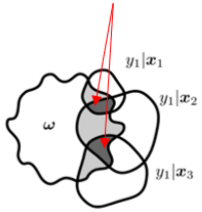


- Find unsafe points in the model (e.g. specific image)
- Query user with specific task (e.g. specific label)
- Update model by user feedback

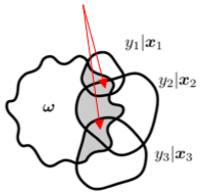
Example: Google ReCAPTCHA [4]



Redundancy:
Scoring single images
Information overlap



Diversity:
Scoring Batches
Information union

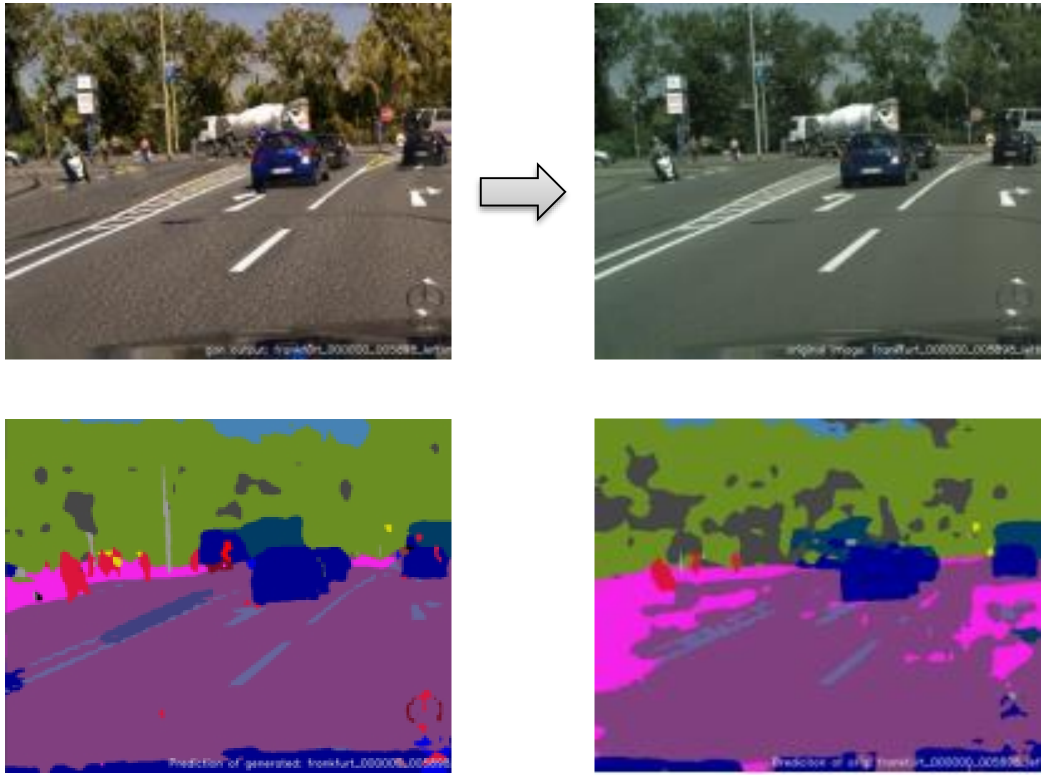


[3] Burr Settles. Active Learning Literature Survey. Computer Sciences Technical Report 1648, University of Wisconsin–Madison. 2009.
[4] Google Developers: ReCAPTCHA. developers.google.com/recaptcha/

Use of synthetic data
Large labeled datasets are cheap now

Straightforward idea

- Generate synthetic data with ground truth, similar to considered real-world domain
  - use high-quality rendering engine (GTA5, CARLA, …)
  - use nearly equal sensor emulation (alignment at vehicle, distortion, curves, radiometry, noise, …)
  - choose similar environment (vehicle path, location, weather, …)
- Learn model ☺
- Possible further steps:
  - active learning: predict real-world labels and let user optimize where prediction is presumably bad
  - optimize model with domain adaptation for best fit with real data



Synthetic World   Real World

Domain Adaptation

How do we learn with new data automatically?

- First of all: we need a labeled source domain with existing AI model
- Then, choose delta-domain to simplify and to increase overlap
  - countries

  - weather

  - sensors

  - long-term change

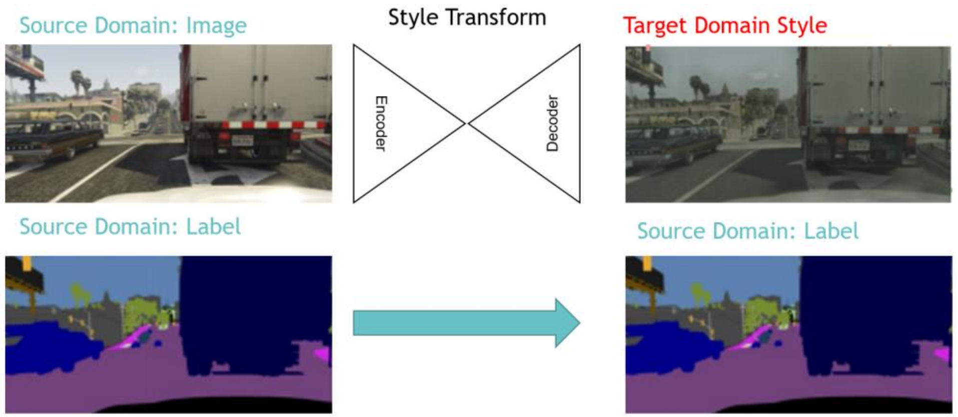
  - season

  - real/synthetic

  - …

- Then, have suitable data in target domain ☺

- Eventually, unsupervised Learning benefits from similarities between domains



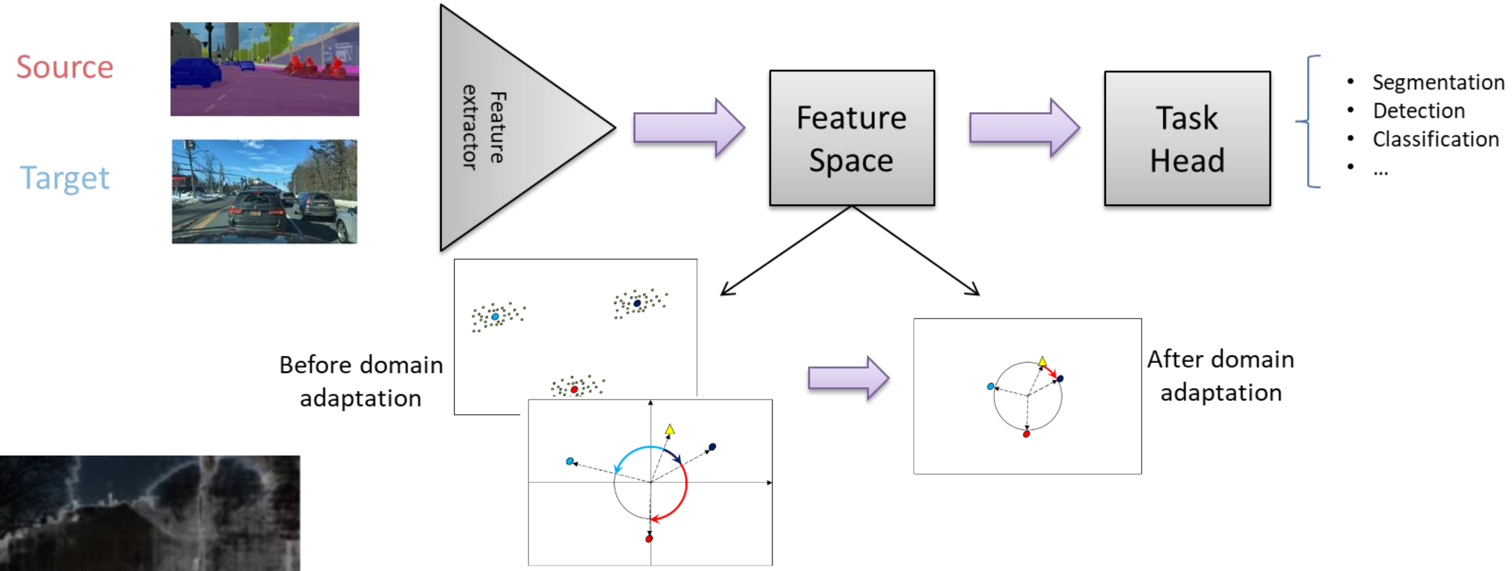
Images from Sefati et al., KI DeltaLearning project presentation - ki-familie.vdali.de

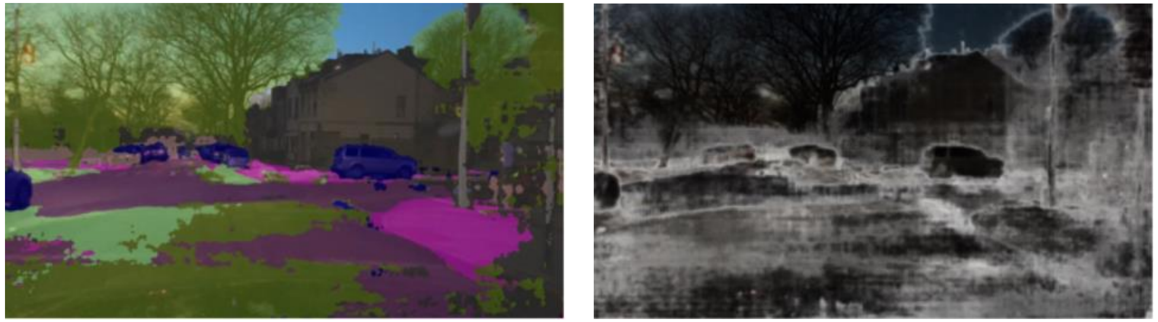
Domain Adaptation

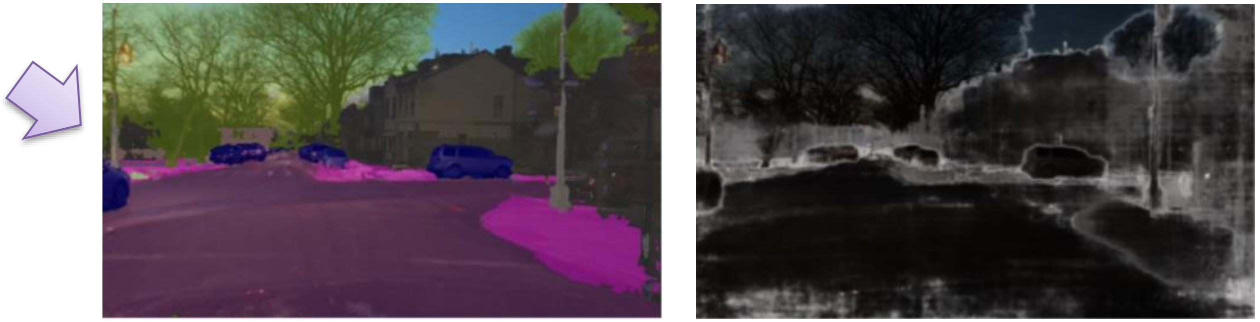
Combine Feature Space to fit with both Domains [5]

Steps:

- align source and target input (see [6])
- decode features in model (e.g. ResNet [7])
- align feature space: minimize entropy in similarity [8] → provides pseudo-labels
- align output space: self-training with pseudo-labels (iterative)







[5] Niemeijer, Schäfer: Combining Semantic Self-Supervision and Self-Training for Domain Adaptation in Semantic Segmentation. Workshop Autonomy@Scale, IV 2021
[6] Hoffman: CyCADA: Cycle-Consistent Adversarial Domain Adaptation, PMLR 2018
[7] He et al.: Deep Residual Learning for Image Recognition, CVPR 2016
[8] Saito et al.: Universal Domain Adaptation through Self Supervision, NeurIPS 2020

Domain adaptation
Result example

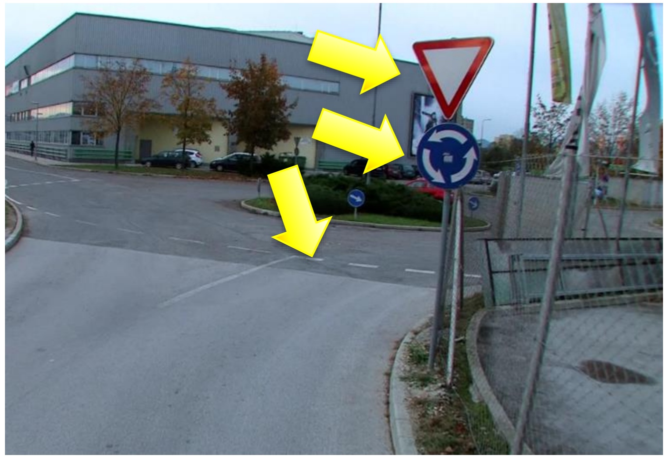
Evaluation on BDD with CityScapes source domain

before domain adaptation

after domain adaptation

What about context?

Limited relevance



Connected signs and markers

Meaning unclear

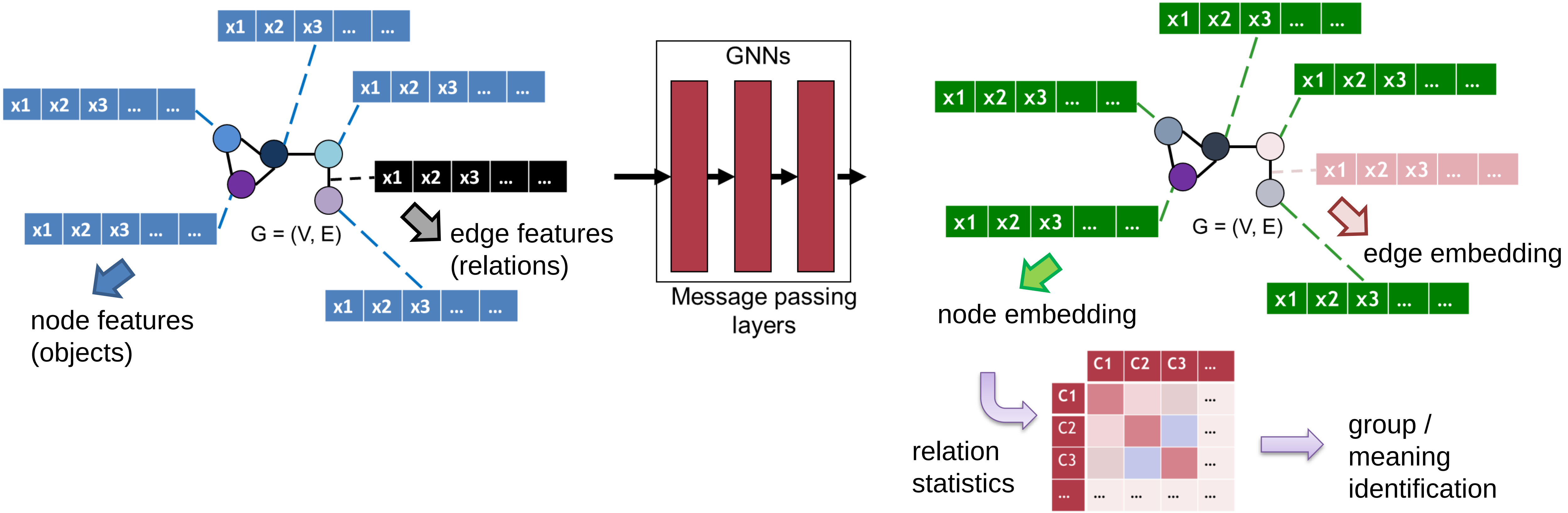
Fakes

Need to know the meaning, not only the type:

- Relations between signs
- Scope and validity of signs
- Relation between signs and road participants
- Expected behavior of road participants
- …

Concept: “Knowledge Directed Object Detector”

Knowledge Representation into Graph Neural Networks [9]

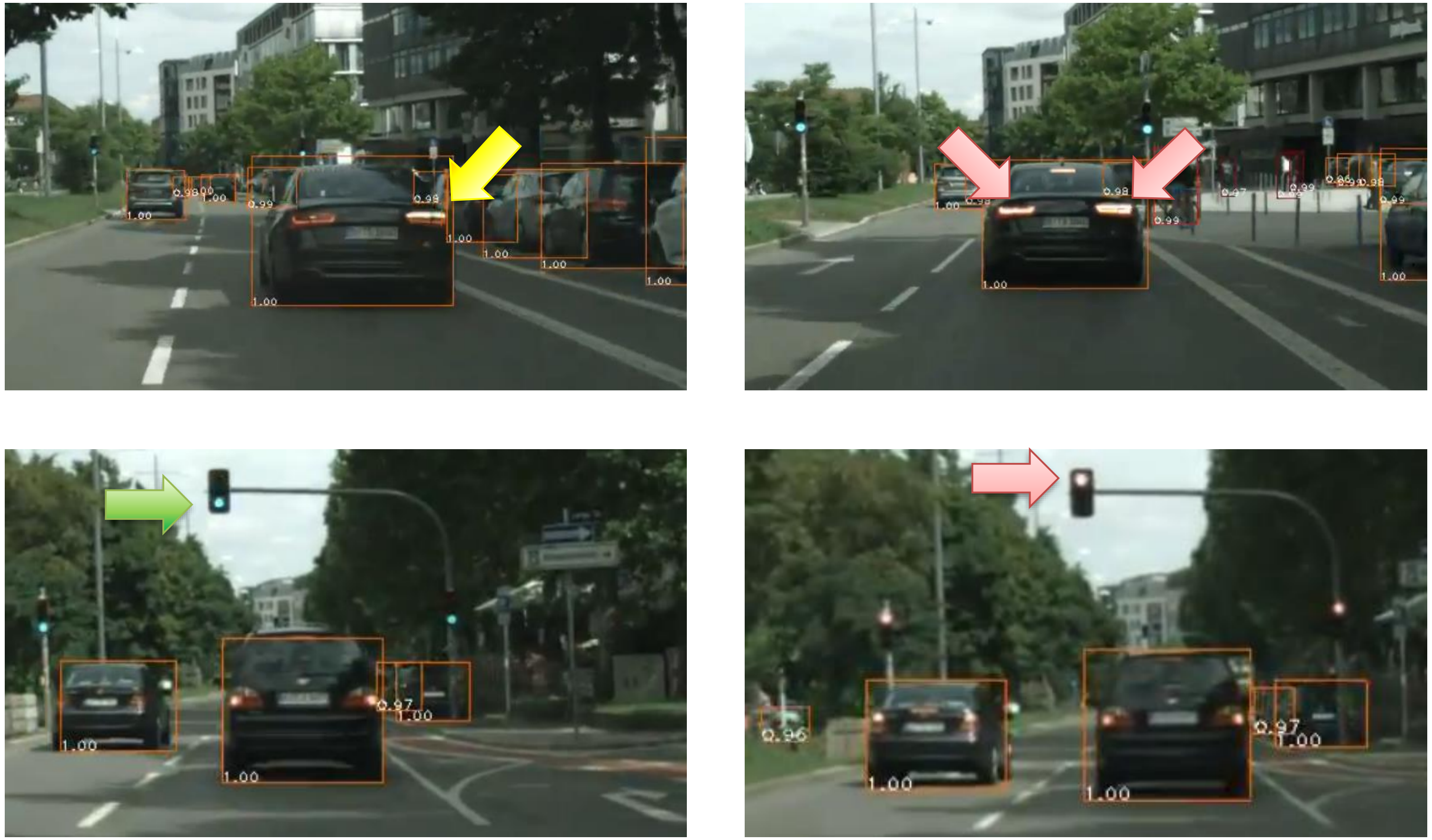


[9] Jiang et al., Hybrid Knowledge Routed Modules for Large-scale Object Detection, NeurIPS 2018

Future work: Prediction of Object Behavior and Risk

What will happen in the next few seconds?

- Old but proven: prediction from filter (e.g. Kalman filter)
  - good at regular driving
  - bad at sharp events **-> Problem!**
- Behavior and movement prediction from indicators
  - at object itself
  - within environment (static / dynamic elements)
  - in mutual interactions
  - in map / rules
  - in meta-data (time, weather, …)
- Indicator labeling and learning
- Video and multi-info-based learning
- Will result in complex ontology of measurable and non-measurable information and knowledge



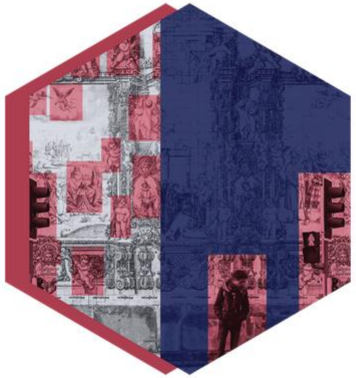
Images from CityScapes Video –
www.youtube.com/watch?v=yNc5N1MOOt4

Thank you!

Project information: ki-familie.vdali.de

KIDELTA LEARNING
Scalable AI for Automated Driving

Methods and tools for the efficient expansion and transformation of existing AI modules in autonomous vehicles to meet the challenges of new domains or more complex scenarios



KI WISSEN
Automotive AI Powered by Knowledge

Development of methods for incorporating knowledge into machine learning

Supported by:

Federal Ministry for Economic Affairs and Climate Action

on the basis of a decision by the German Bundestag

Contact: Dr.-Ing. Franz Andert
DLR-Institute of Transportation Systems
franz.andert@dlr.de

DLR