KI DATA TOOLING
The Data Kit for Automotive AI

KI Data Tooling Final Event | 05/06 December 2023

Deep Dive #1
Training with Synthetic Data - Mixed Training

Maarten Bieshaar, Maximilian Menke, Robert Bosch GmbH

Training with Synthetic Data Team @ Bosch


Maarten
Bieshaar


Cheuk Yan
Law


Christoph
Dreißigacker


Simon
Heming


Emad
Ibrahim


Armin
Köhler


Patrick
Heer


Maximilian
Menke


Falko
Matern

Introduction
Mixed Training - A Paradigm Shift

Goal:

Integrate Synthetic data into Data-Driven Engineering Process [1]

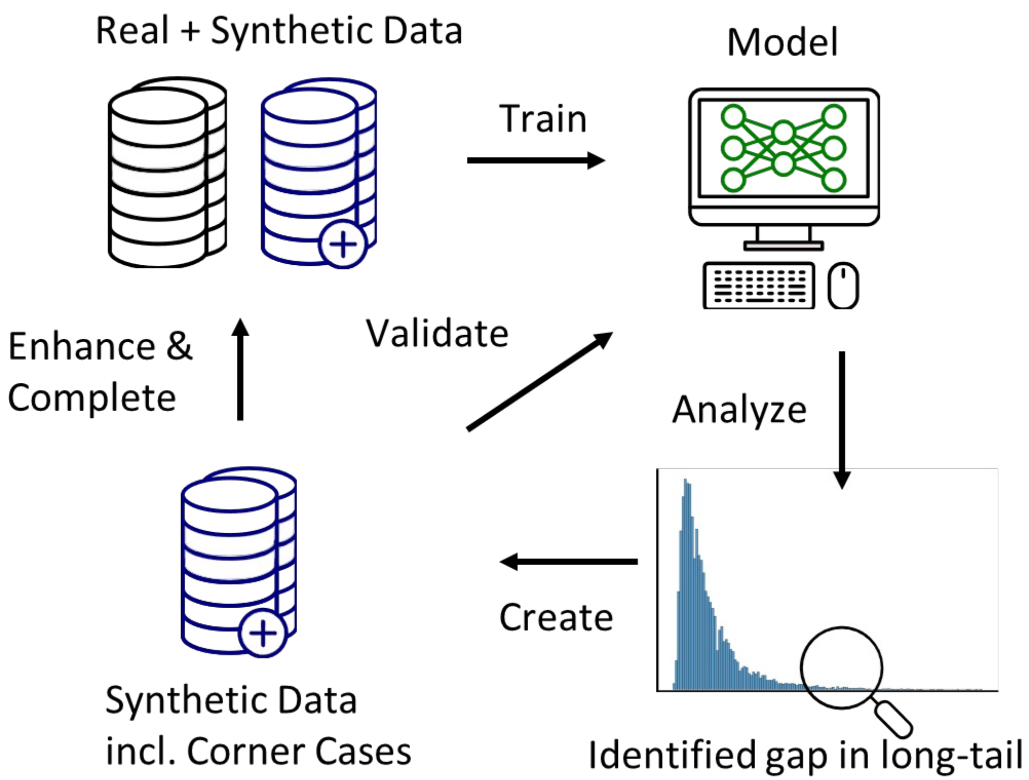
Paradigm Shift:
Generate the data the model really needs!

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

Research Questions:

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

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



1

Mixed Training Challenge

Mixed Training Challenge
Let's construct a gap

Goal of challenge:

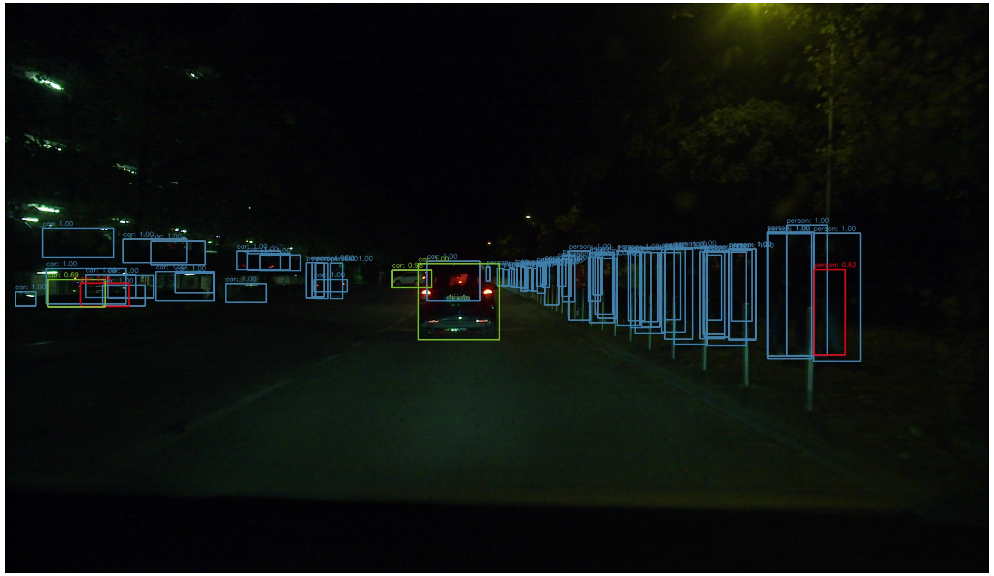
Use synthetic data to compensate for missing data in real training dataset

Night Images:

Challenging due to challenging lighting conditions and groups of pedestrians

Artificial gap creation:

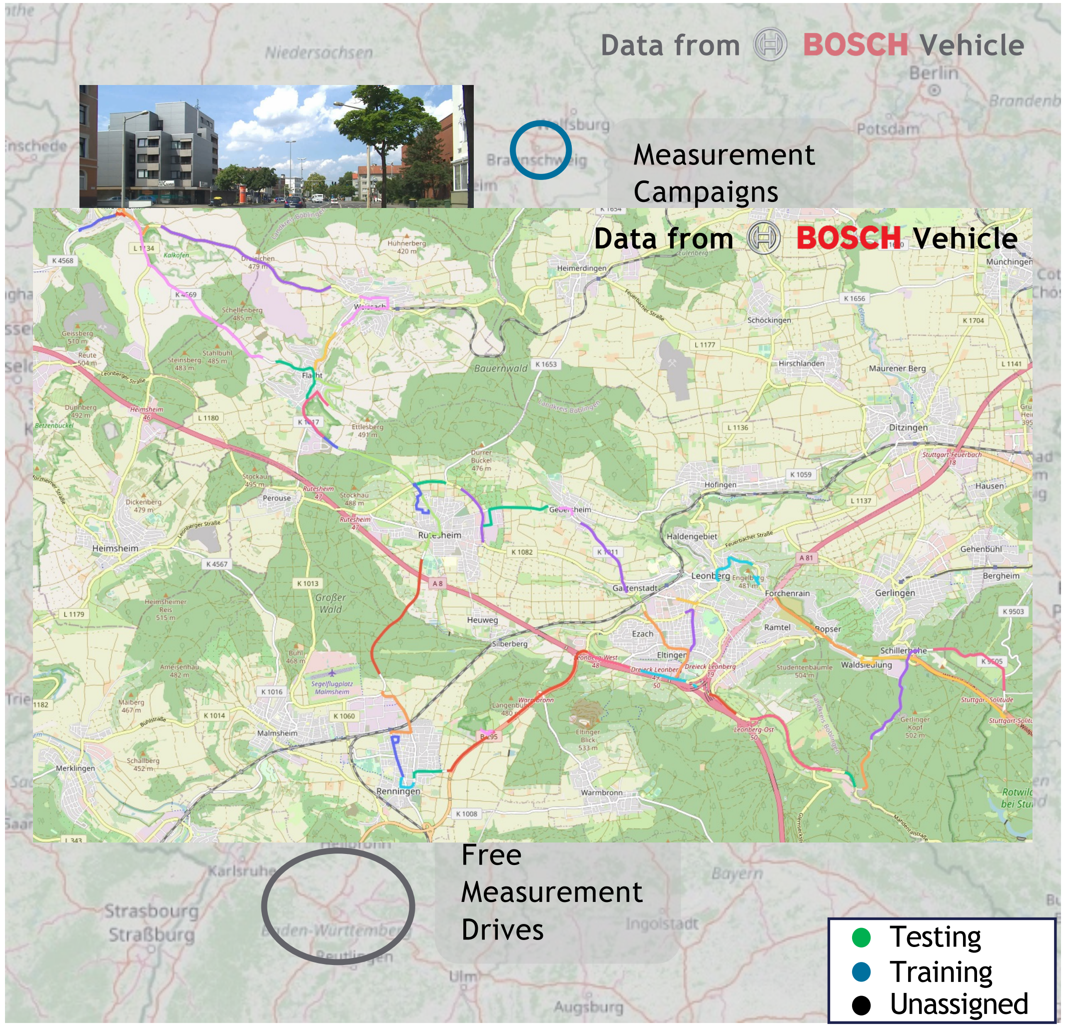
Remove night images from real training data



→ Synthetic data should be used to improve pedestrian detection at night!

Mixed Training Challenge
Data splits overview

- Training and Test set **independence** via
  - **Spatial:** Frames in splits are recorded in different locations
  - **Temporal:** Split on the level of sequences



Mixed Training Challenge
Dataset Overview

	Synthetic Data (SynPeDS [2])	Real data (Bosch Vehicle @ KI-DT)
Camera, Image Resolution (HxW)	Pinhole, 1920 x 1280	Bosch Atom Camera, 1920 x 1080
# Labeled Images	123.721	13.650
Location	Synthetic Urban Areas (Germany)	Urban as well as rural areas (Germany).
Features	Different time of day (e.g., day, night), weather conditions (e.g., dry, wet, and foggy, sun elevations (e.g., low, high), groups of pedestrians, occlusion	Different time of day (e.g., dawn, day, night), weather conditions (e.g., dry, rainy, snowy), sun elevations (e.g., low, high), road types (e.g., country road, highway), groups of pedestrian
Sample images		 

[2] Stauner, T., et al, "SynPeDS: A Synthetic Dataset for Pedestrian Detection In Urban Traffic Scenes," in Proceedings of the 6th ACM Computer Science in Cars Symposium, 2022.

Mixed Training Challenge
Evaluation and Metrics

- Performance evaluation via Log-Average Miss-Rate (LAMR)
- Different Evaluation protocols, i.e., Reasonable, Small, Occluded, All (cf. CityPersons [3])

Problem: Model architecture might inherently skew or bias the result

- Relative performance improvement (avoids skewness due to different model architectures etc.)

$$\left(1 - \frac{LAMR^{Mixed}_{Reasonable\,@\,test\_all}}{LAMR^{Real\ All}_{Reasonable\,@\,test\_all}}\right)\times 100\ \%$$

Performance @ Night:

- Relative model performance improvement for night images
- Further evaluation protocols: Fair, Moderate, and Hard.

Mixed Training Performance:

$$LAMR^{Mixed}_{Reasonable\,@\,test\_all} = 20$$

Real Baseline:

$$LAMR^{Real\ All}_{Reasonable\,@\,test\_all} = 25$$

Relative Improvement Score:

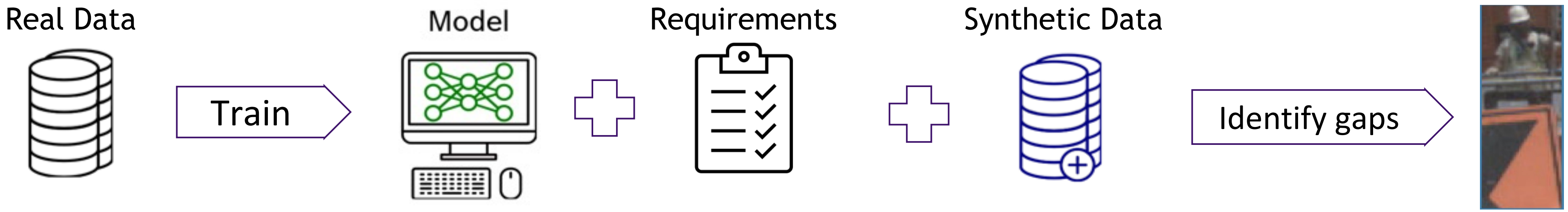
$$\left(1 - \frac{20}{25}\right)\times 100\ \% = 20\ \%$$

[3] S. Zhang, R. Benenson, B. Schiele, "CityPersons: A Diverse Dataset for Pedestrian Detection," in CVPR, 2017.

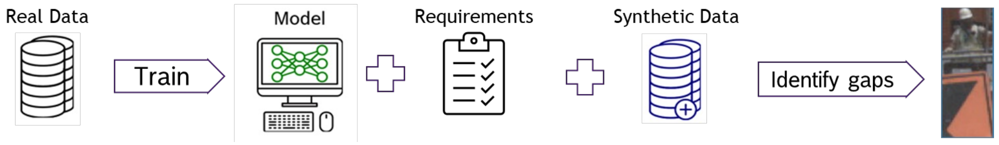
2

Identification of Gaps

Identification of Gaps
Checking Model Performance against Requirements



Identification of Gaps
Simple Case Study



Research Question: Can we use synthetic data to identify these gaps?

Approach:

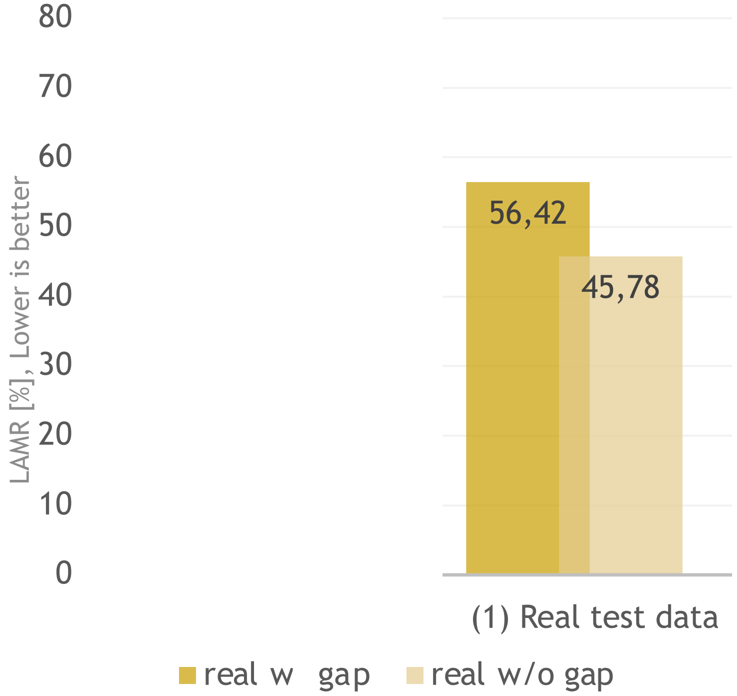
- *Starting point*: Trained model on real data containing an artificially generated gap
- *Study real and synthetic data from the gap:*
  "Check if the model's performance on synthetic test data relates to the performance on real test data in the gap."

Example Use Case:

- Height of objects, (i.e., remove 90 % of boxes with height 34px – 51px)

Can use synthetic data to identify data gaps.
(at least for this simple case study)

Is a real data gap visible in synthetic data?



Train on real data with (■) and w/o gap (■) ,
Evaluate on (1) real and (2) synthetic data in gap

3

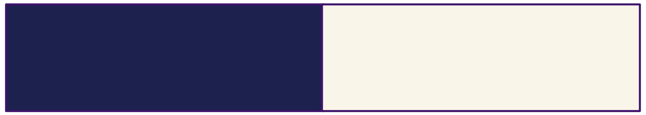
Training Strategies

Training Strategies
Mixing

Research Question: Which mixing strategy to use?

Dataset-Level Mixing

Batch-Level Mixing



Pre-Training → Fine-Tuning (PF-FT)

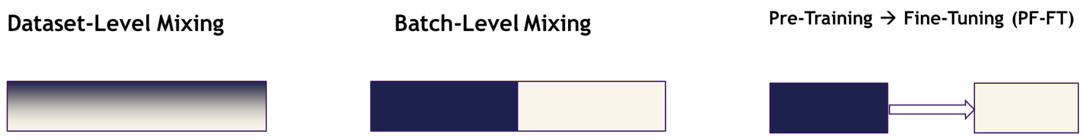
synthetic data

real data

Experimental Setup:

- RetinaNet with ResNet50 Backbone + Feature-Pyramid Network
- AdamW Optimizer
- Use augmentions such as random scaling, rotation, brightness variation etc.

Training Strategies
Mixing



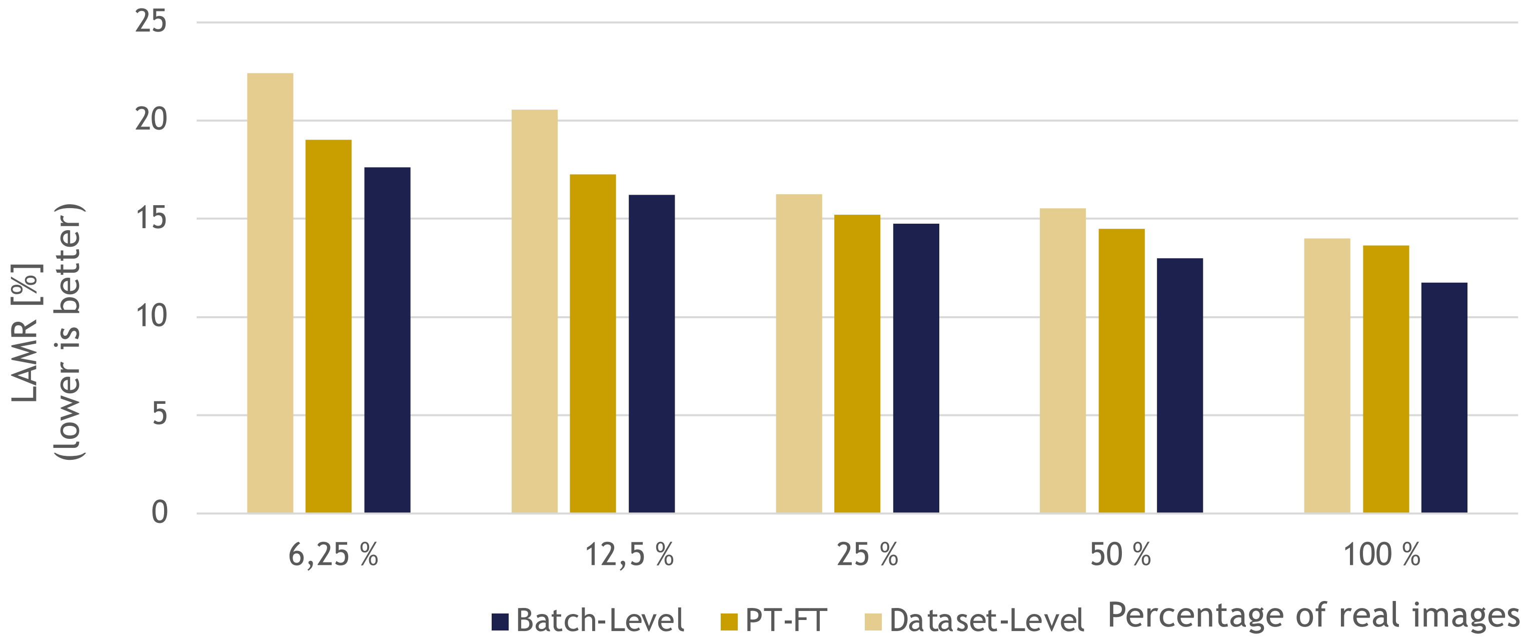
Experimental Setup:

- Datasets:

SynPeDS [2]

CityPersons [3]

- Fixed ratio of real data
  (5 % real, and 95 % synthetic data)



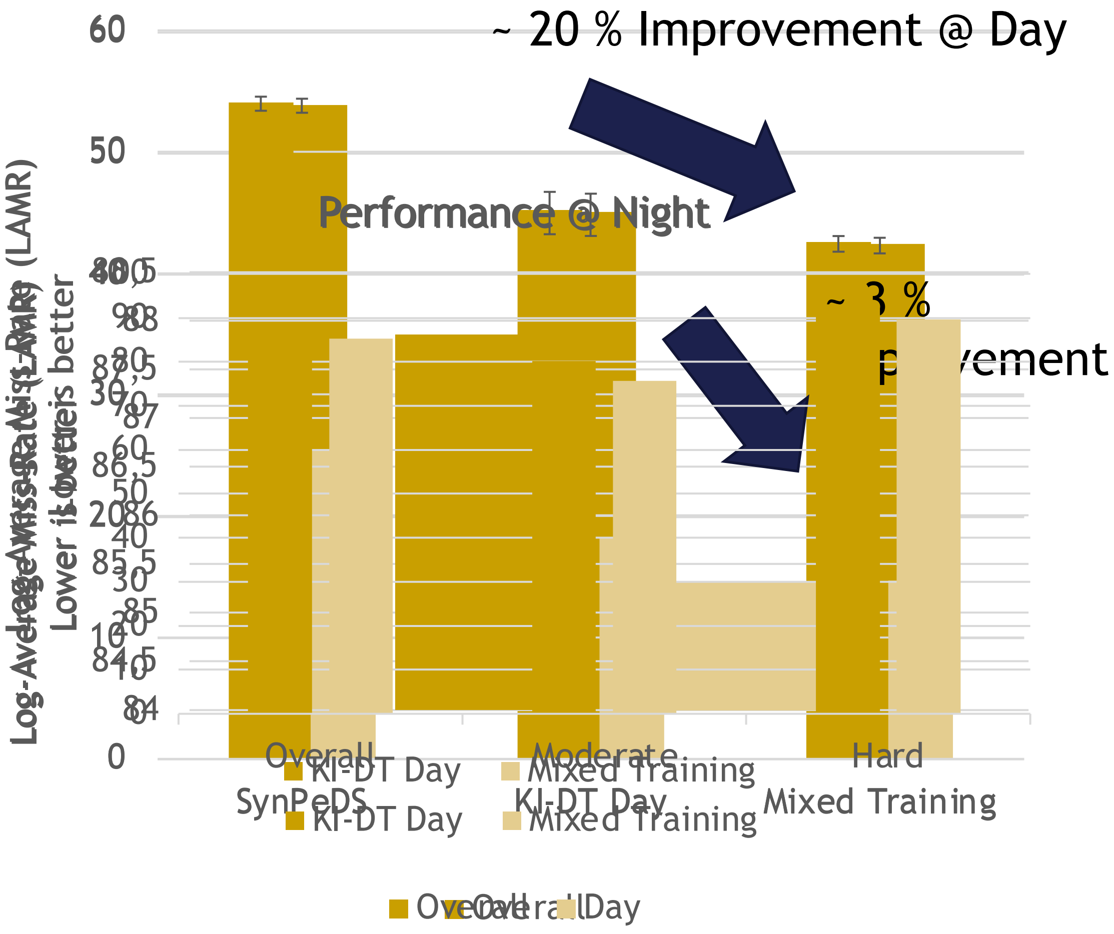
➡ Mixing on the level of batches, i.e., the two-urn model, works best!

[2] Stauner, T., et al, "SynPeDS: A Synthetic Dataset for Pedestrian Detection In Urban Traffic Scenes," in Proceedings of the 6th ACM Computer Science in Cars Symposium, 2022.
[3] S. Zhang, R. Benenson, B. Schiele, "CityPersons: A Diverse Dataset for Pedestrian Detection," in CVPR, 2017.

Training Strategies
KI-DT Challenge - Results

- Batch-level mixing on KI-DT and SynPeDS
- 3 repetitions of each experiments

- Key Results:
  - **Mixing improves model performance**
  - **Better detection at night**



Results on KI-DT test split

4

Image Stylization

Image Stylization
Typical GAN-based Style-Transfer

Major challenge and performance limitation:

- **Domain Gap** between real and synthetic data.

Image-to-Image style transfer to mitigate gap

- **First stage** of the mixed training
- GAN-based input-level domain adaptation

Detector Training and Inference

- **Second stage** of the mixed training
- Object Detection training on stylized data
- Inference on real-world target domain data

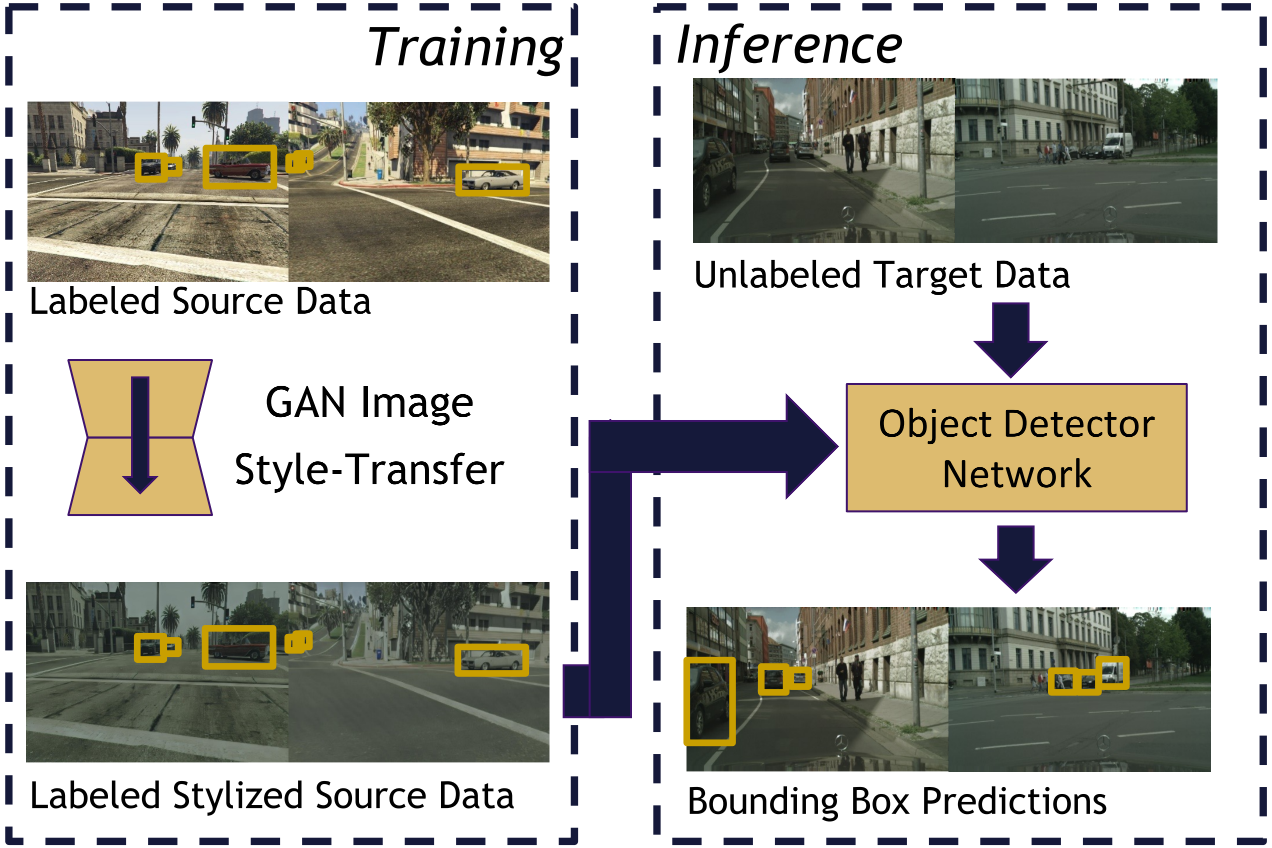
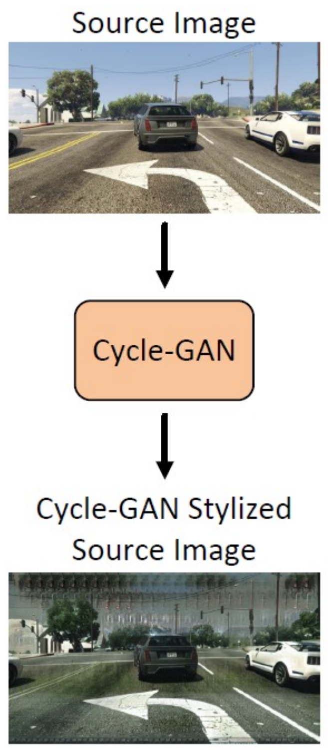


Image Stylization
AWADA Style-Transfer

a) **Cycle-GAN** [4] Style-Transfer is not optimal for detection due to hallucinations

b) **AWADA** [5] Style-Transfer optimized for object detection by using attention maps separating foreground and background in loss optimization



(a)

[4] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks." Proceedings of the IEEE international conference on computer vision. 2017.
[5] Menke, Maximilian and Wenzel, Thomas and Schwung, Andreas, Awada: Foreground-Focused Adversarial Learning for Cross-Domain Object Detection. Available at SSRN: https://ssrn.com/abstract=4272713 or http://dx.doi.org/10.2139/ssrn.4272713

Image Stylization
AWADA Style-Transfer

Original GTA Image

Cycle-GAN Stylized GTA Image

AWADA Stylized GTA Image

Image Stylization
AWADA Qualitative Results

Cityscapes-to-FoggyCityscapes

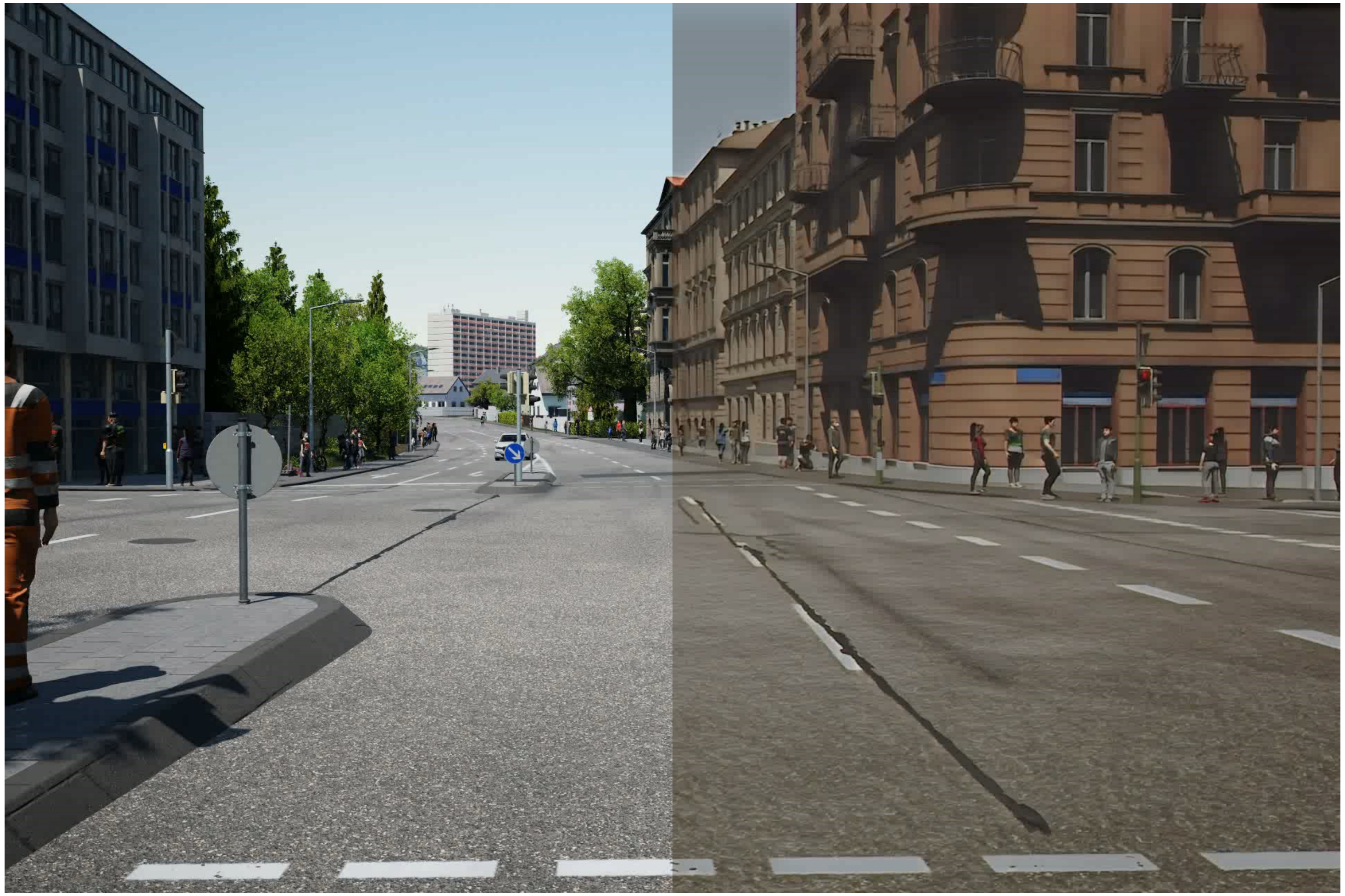
GTA5-to-Cityscapes

Cityscapes-to-BDD100k

Image Stylization
AWADA Qualitative Results

KIA

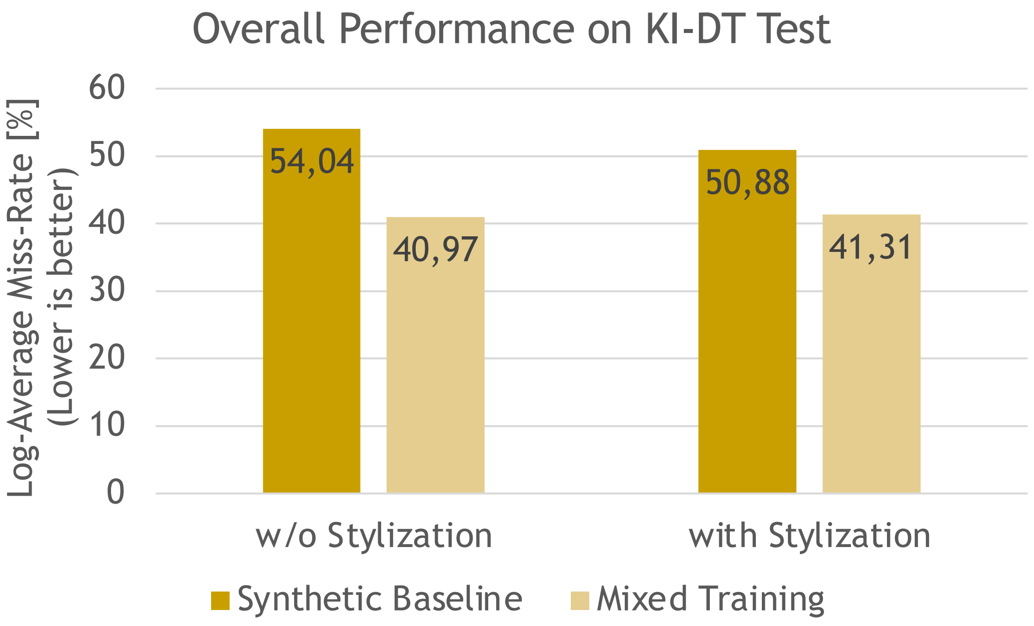
Dataset

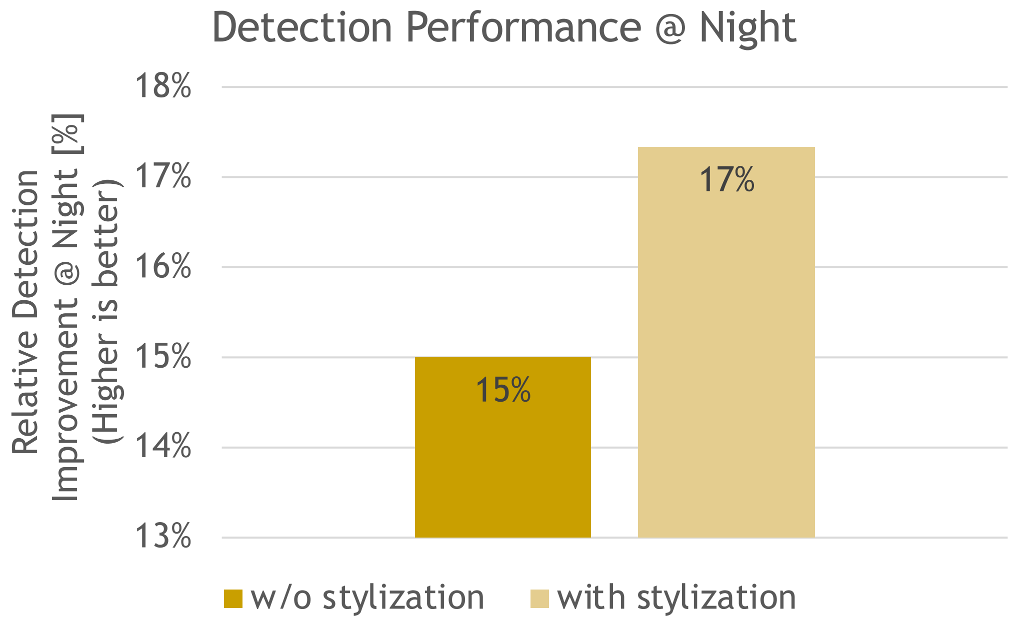


KIA-to-KIDT

Dataset

Image Stylization
KI-DT Challenge - Results





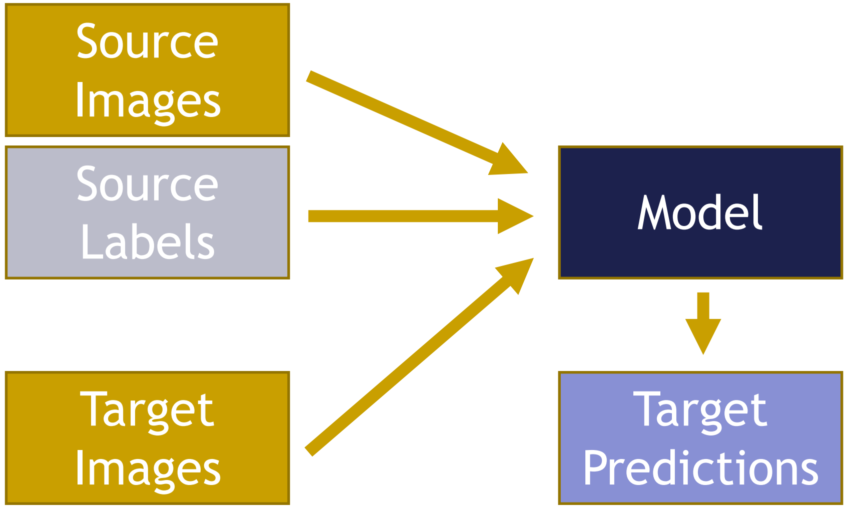
- **Stylization helps for synthetic data pre-training not for joint mixed training**

5

Unsupervised Domain Adaptation

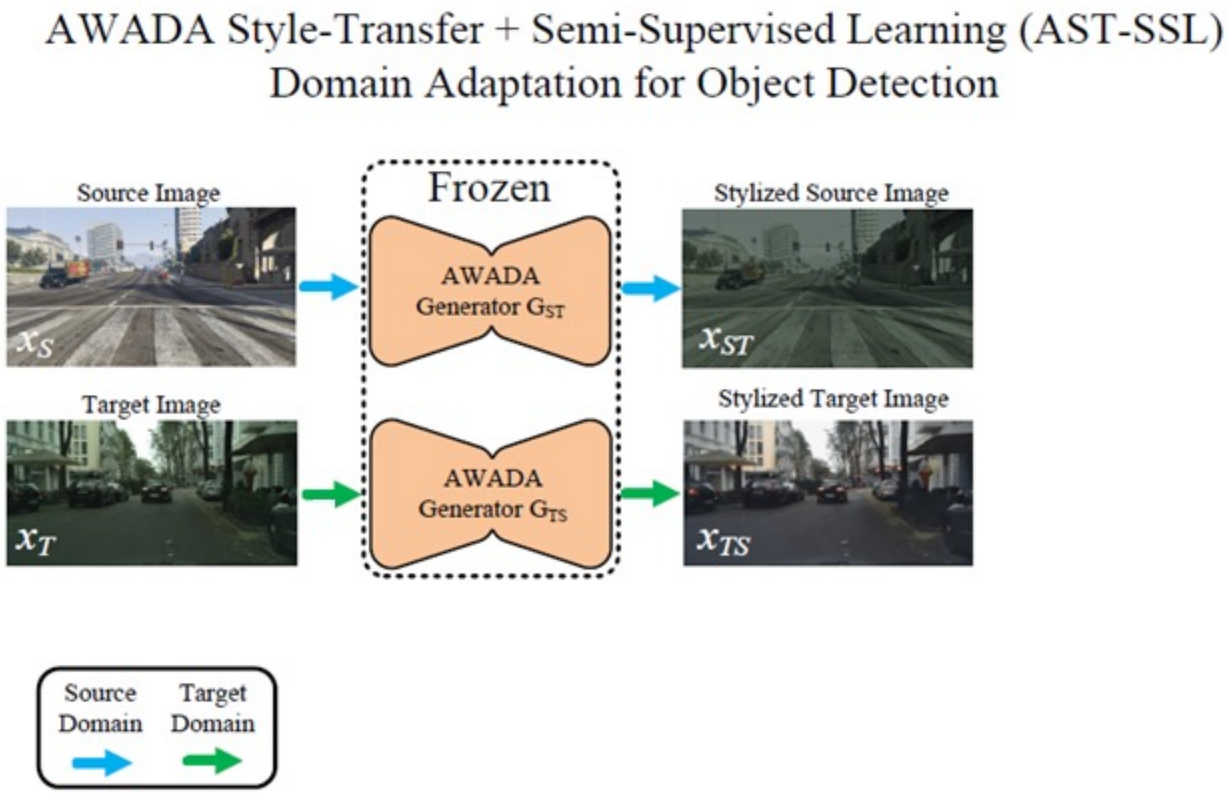
Unsupervised Domain Adaptation
Introduction

- Unsupervised Domain Adaptation (UDA)
  - **Supervision** from the source domain (synthetic data)
  - **No supervision** from the target domain (real data)
- When to apply UDA?
  - Image + labels from the source domain
  - Unlabeled images from the target domain



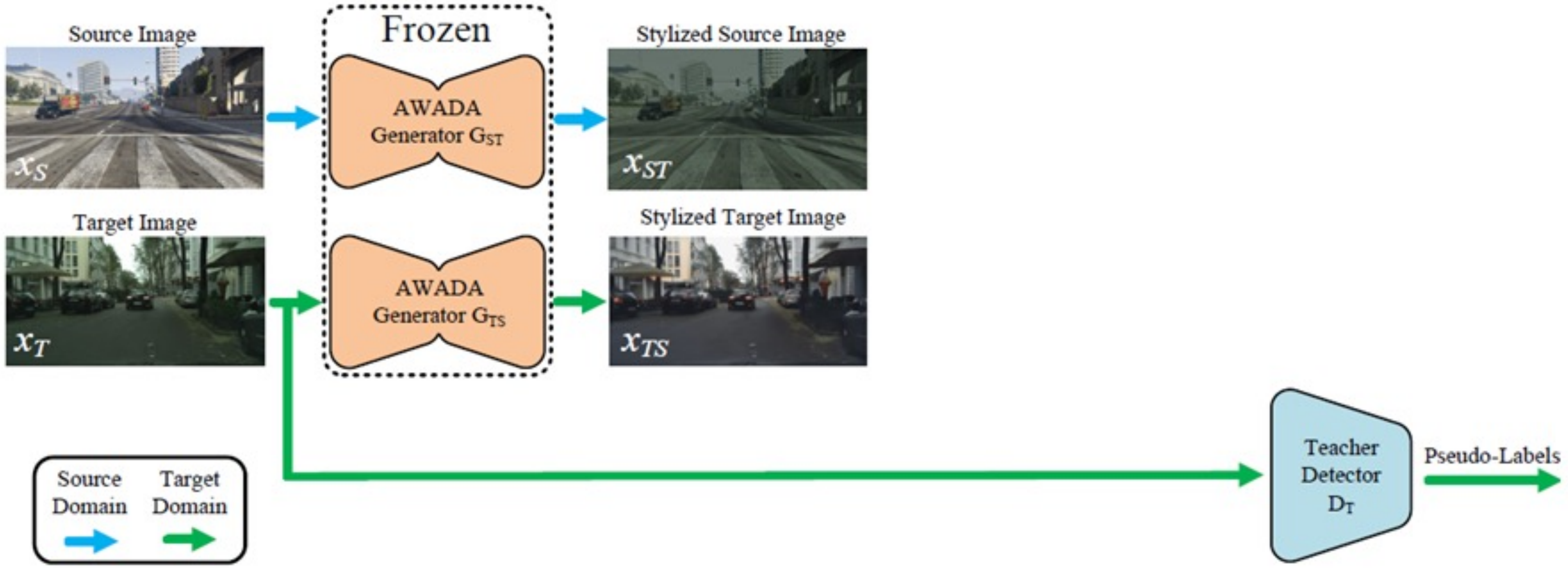
- Real use-cases of the UDA setting:
  - Target domain data is **hard and costly to annotate** (Semantic Segmentation / Object Detection)
  - **Labeling is in process**, but training on unlabeled target domain data can already be started

Unsupervised Domain Adaptation
Semi-Supervised Learning Framework



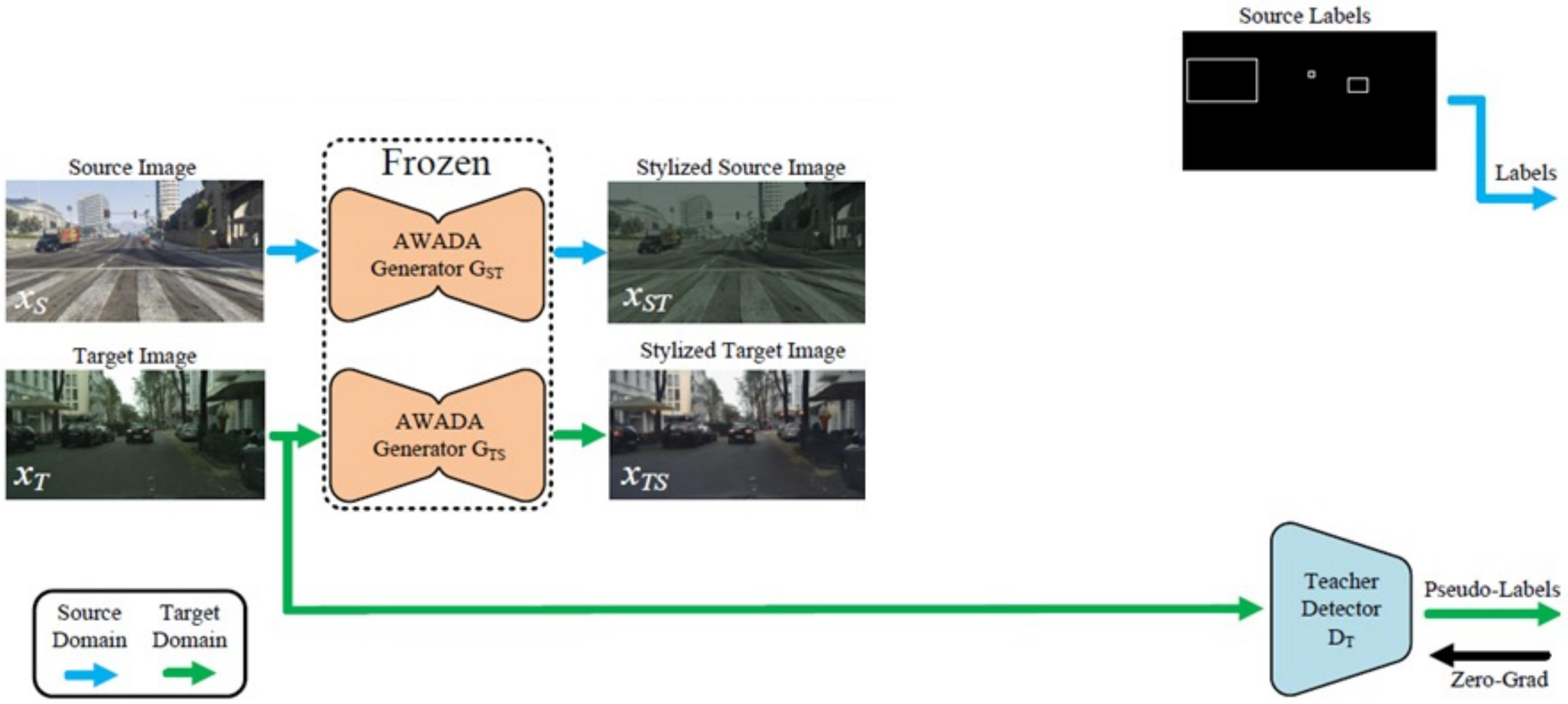
[6] Menke, Maximilian, Thomas Wenzel, and Andreas Schwung. "Improving Cross-Domain Semi-Supervised Object Detection with Adversarial Domain Adaptation." 2023 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2023.

Unsupervised Domain Adaptation
Semi-Supervised Learning Framework



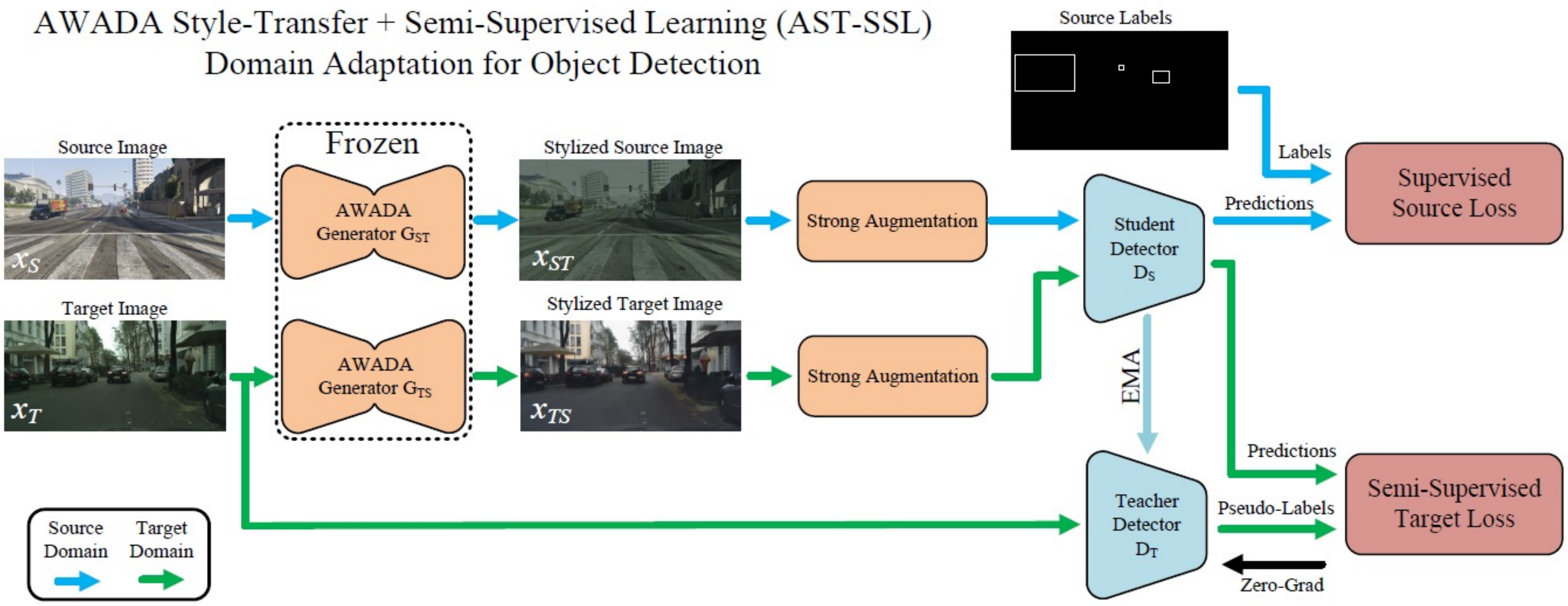
[6] Menke, Maximilian, Thomas Wenzel, and Andreas Schwung. "Improving Cross-Domain Semi-Supervised Object Detection with Adversarial Domain Adaptation." 2023 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2023.

Unsupervised Domain Adaptation
Semi-Supervised Learning Framework



[6] Menke, Maximilian, Thomas Wenzel, and Andreas Schwung. "Improving Cross-Domain Semi-Supervised Object Detection with Adversarial Domain Adaptation." 2023 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2023.

Unsupervised Domain Adaptation
Semi-Supervised Learning Framework



[6] Menke, Maximilian, Thomas Wenzel, and Andreas Schwung. "Improving Cross-Domain Semi-Supervised Object Detection with Adversarial Domain Adaptation." 2023 IEEE Intelligent Vehicles Symposium (IV). IEEE, 2023.

Unsupervised Domain Adaptation
Results on Public Datasets

- Faster R-CNN object detector with VGG16 Backbone
- Mean Average Precision (mAP) @ 0.5 IoU Threshold (Higher is better)

Method	GTA-to-Cityscapes (Synth-to-Real) [%]	Cityscapes-to-FoggyCityscapes (Real-to-Foggy) [%]	Cityscapes-to-BDD100k (Real-to-Real) [%]
Synthetic Baseline	35.7	29.3	25.7
SOTA [7]	53.1	50.9	29.6
AWADA	54.1	44.8	31.5
SSL + AWADA (AST-SSL)	**57.3 (+3.2)***	**51.5 (+0.6)***	**41.6 (+10.1)***
Real Baseline (Oracle)	70.0	47.9	53.4

- **Result:** AST-SSL new SOTA in unsupervised domain adaptation for object detection.

* Relative improvements w.r.t. AWADA.

[7] Khindkar, Vaishnavi, et al. "To miss-attend is to misalign! residual self-attentive feature alignment for adapting object detectors." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2022.

Unsupervised Domain Adaptation
Results on KI-A to KI-DT

- RetinaNet Object Detector with ResNet50 Backbone

Method	LAMR (overall) [%]	LAMR (day) [%]	LAMR (night) [%]
Synthetic Baseline	58.5	30.5	95.2
AWADA	58.1 (-0.4)	29.9 (-0.6)	94.4 (-0.8)
SSL	**54.9 (-3.6)**	**26.5 (-4.0)**	**94.2 (-1.0)**
SSL + AWADA	57.9 (-0.6)	29.6 (-0.9)	95.2 (+0.0)
Real Baseline (Oracle)	50.9	24.8	93.4

- **AST-SSL considerable improves**
- **Challenge:** Results of AST-SSL (current SOTA) differ to those obtained on public benchmarks
  - Current benchmarks are too easy,
  - Pedestrian detection at night not improved so far.

6

Conclusion

Conclusion
Summary

- **Identification of Gaps: No need for perfect synthetic data**
  - We can already prove performance deficits with "simple" synthetic data
- **Mixing Strategies: Joint training on synthetic and real data**
  - Mixing on level of batches works best, but you'd also better check-out image-level mixing.
- **Domain Adaptation: Style Transfer**
  - Improves visual realism of synthetic data and downstream model performance.
- **Unsupervised Domain Adaptation: Reducing Annotation Effort**
  - 91% of the performance of fully supervised model w/o any real labeled data.
- **Mixed Training Challenge: Improve comparability and foster research on mixed training**
  - Synergetic combination of synthetic and real data for targeted filling of gaps
  - Upload and publication aspired, stay tuned!

KI DATA TOOLING
The Data Kit for Automotive AI

Maarten Bieshaar, maarten.bieshaar@de.bosch.com

Maximilian Menke, maximilian.menke@de.bosch.com

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

KI FAMILIE

www.ki-datatooling.de @KI_Familie KI Familie

Supported by:

Federal Ministry for Economic Affairs and Climate Action

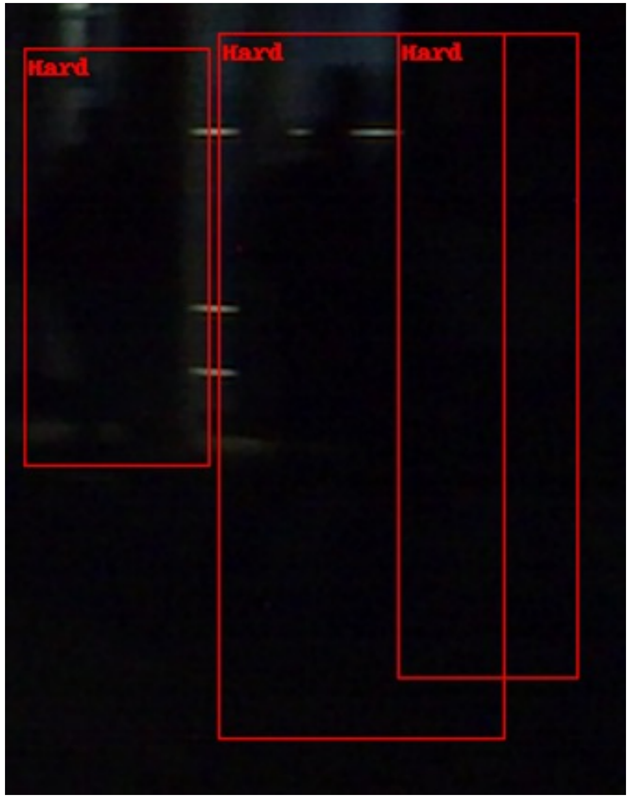
on the basis of a decision by the German Bundestag

7

Backup

KI-DT Mixed Training Challenge
Difficulty of Night Images





KI-DT Mixed Training Challenge
Difficulty of Night Images