Deep-Learning-Based Sensor Fusion for 3D Object Detection with Camera and Low-Cost LiDAR Sensors

14.07.2022, Dr. Cay-Christian Kalmbach

bertrandt

01 Introduction - 3D Object Detection

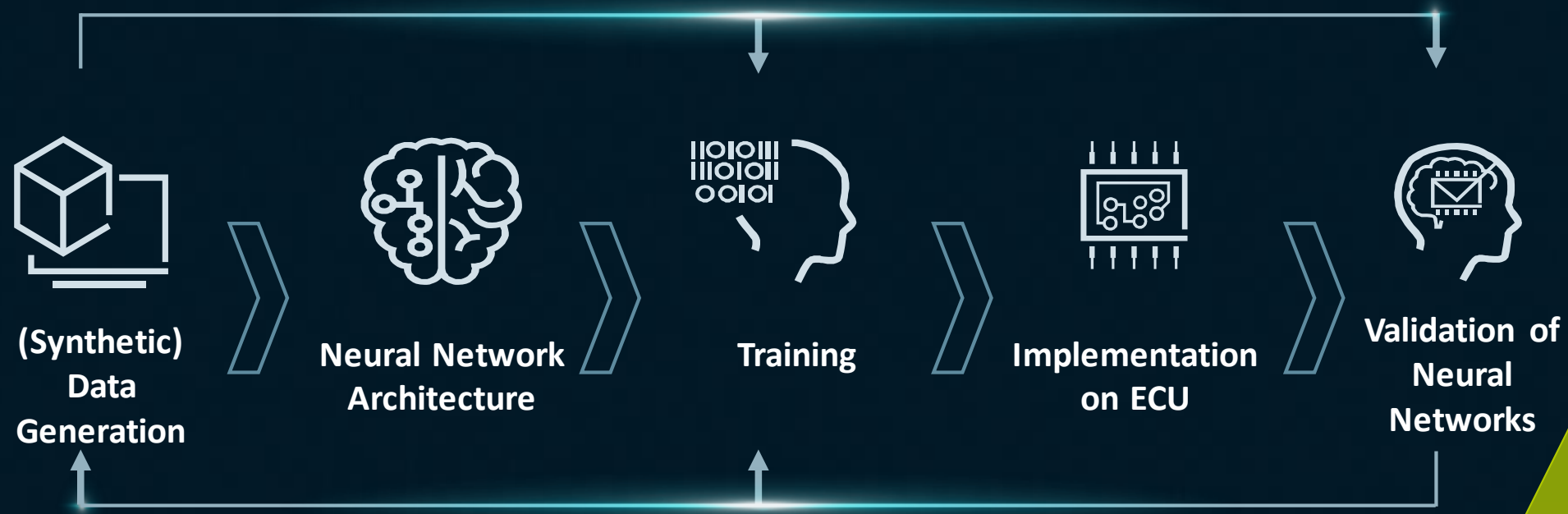
02 AI-Based Sensor Fusion

03 Our Approach: PSR-PointNets

04 Synthetic Data for AI Development & Validation

05 Summary

Deep Learning Workflow



In this talk: results from a publicly-funded research project

3D Object Detection with Camera and LiDAR Data

bertrandt

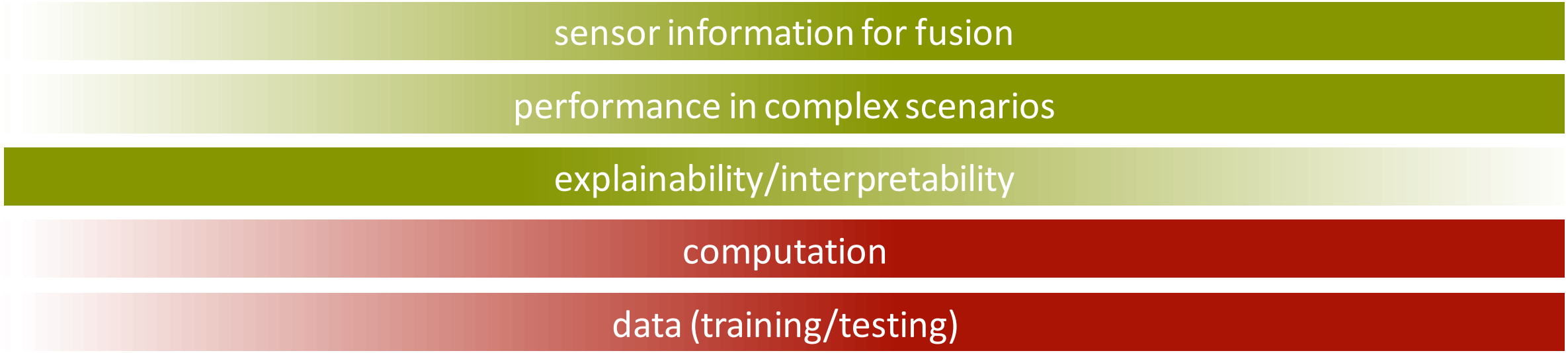
Sensor Fusion: From Object-Level to AI-Based Feature Level Fusion

Currently: Object-Level Fusion

- High-level fusion on object-level or track-level
- Classical algorithms (clustering, Kalman, etc.)
- Which sensor to trust if they contradict?

Trend: AI-Based Feature-Level Fusion

- Mid-/low-level fusion on feature level
- Deep learning algorithms
- Learns to weight the sensor information



Sensor Configuration – Camera & Low-Resolution LiDAR

In our experiments:
modified KITTI dataset
(bottom)

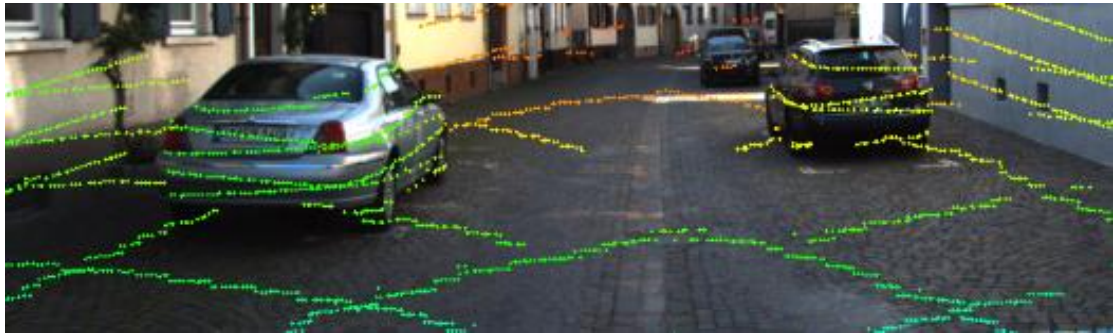
Source (KITTI dataset): A. Geiger, P. Lenz, C. Stiller and R. Urtasun, Vision meets Robotics: The KITTI Dataset, International Journal of Robotics Research (IJRR), 2013

Basic Ideas of our Approach

The **image + one LiDAR point** on the object **is enough** → a human can annotate a good 3D bounding box

- The **(low-resolution) measured** LiDAR point cloud **is enriched** with a **(high-resolution) image-predicted** object point cloud
- The neural network …
  - > … **predicts the high-resolution point cloud** from image information.
  - > … is trained with high-resolution LiDAR data …
  - > … but **in the car, only a low-resolution LiDAR** is needed

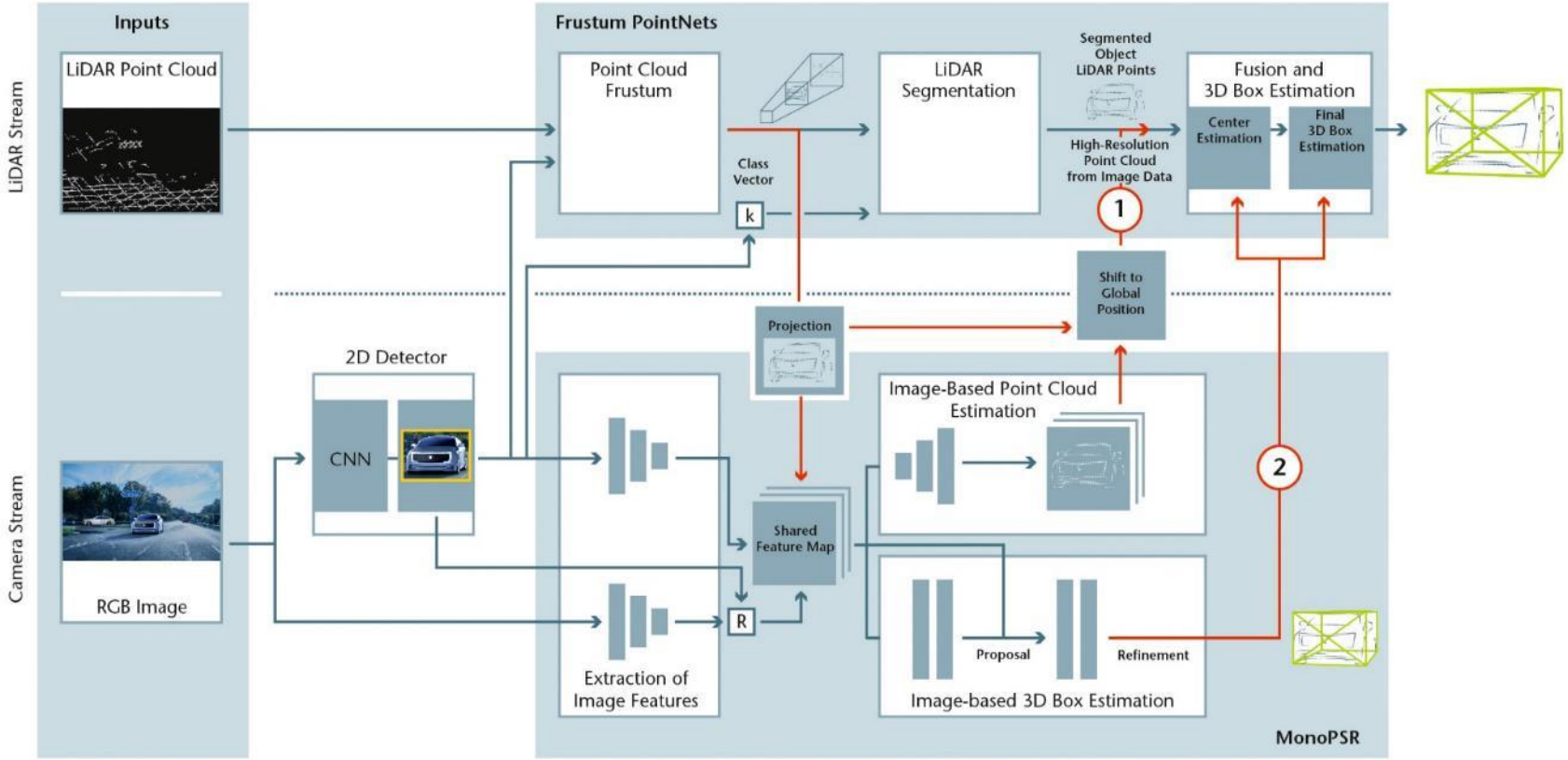
Low-resolution measured point cloud of the object



High-resolution image-predicted point cloud of the object

Source (KITTI dataset): A. Geiger, P. Lenz, C. Stiller and R. Urtasun, Vision meets Robotics: The KITTI Dataset, International Journal of Robotics Research (IJRR), 2013

Our Approach: PSR-PointNets

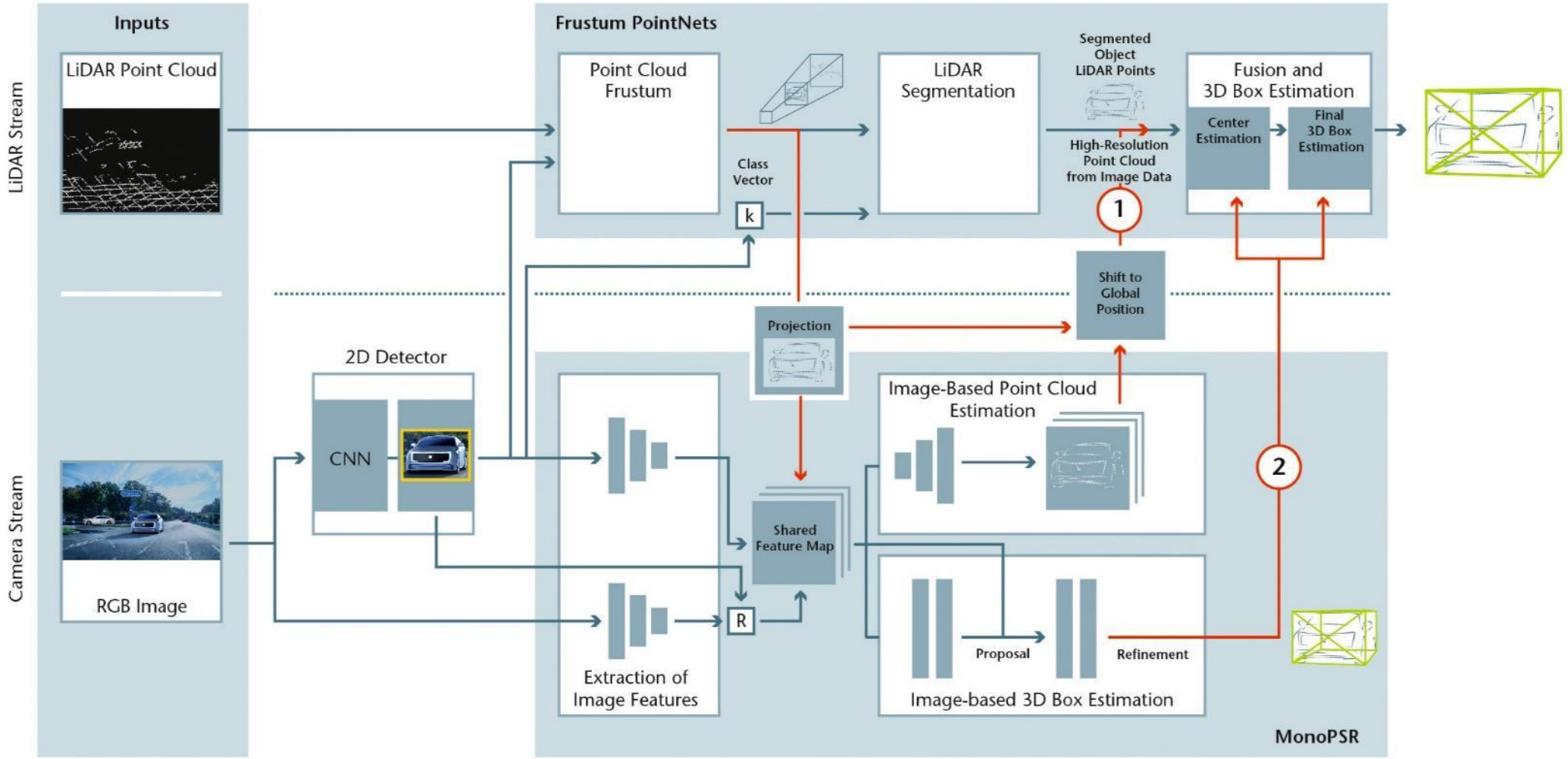


3D Object Detection

- Camera + low-cost LiDAR sensors
- AI-based sensor fusion
- Fusion of 3D reconstruction of the object from image data with measured LiDAR data
- Improved explainability:
  - > Modular architecture
  - > Human-interpretable intermediate results

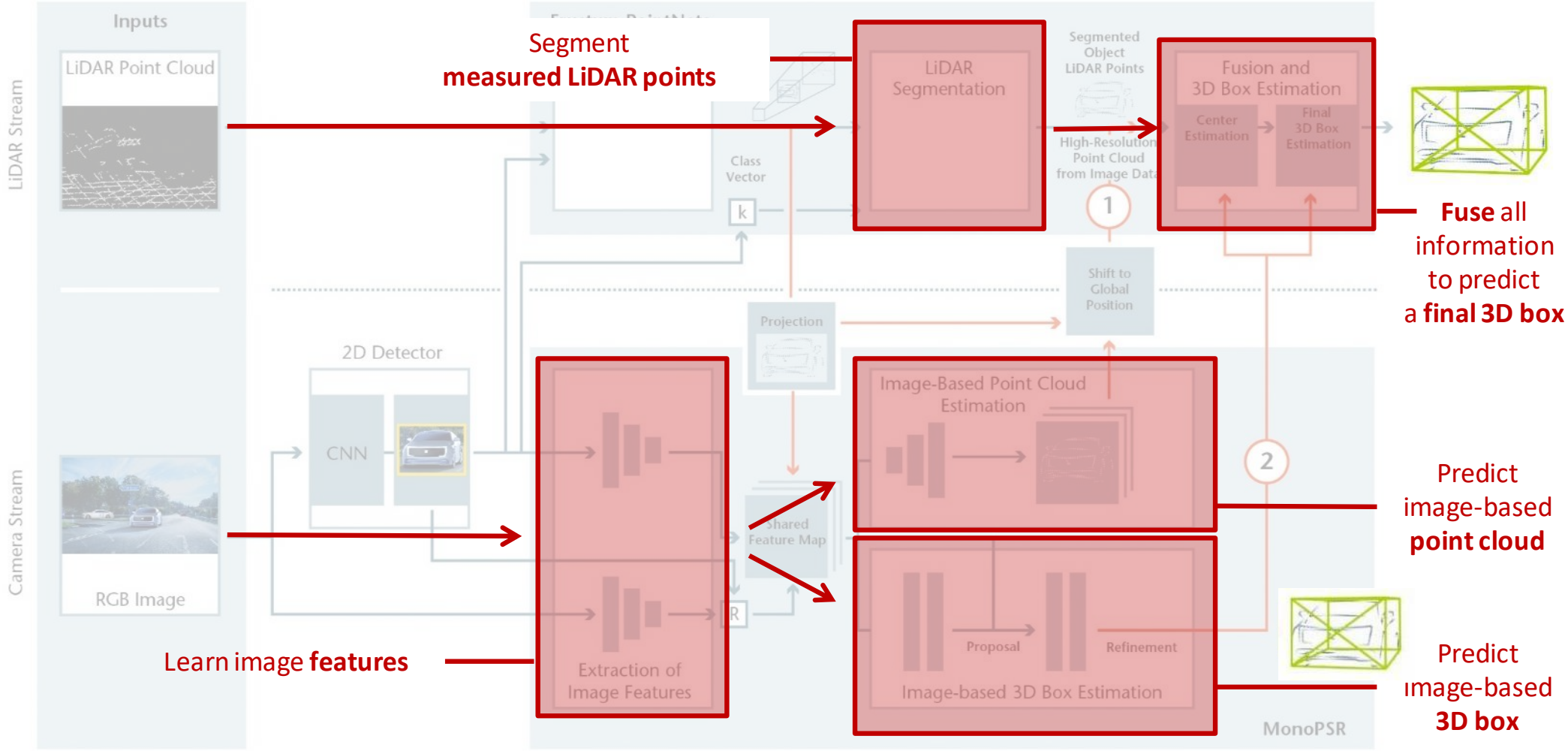
Source (Frustum PointNets): C. R. Qi et al., https://arxiv.org/abs/1711.08488
Source (MonoPSR): J. Ku et al., https://arxiv.org/abs/1904.01690

Our Approach: PSR-PointNets



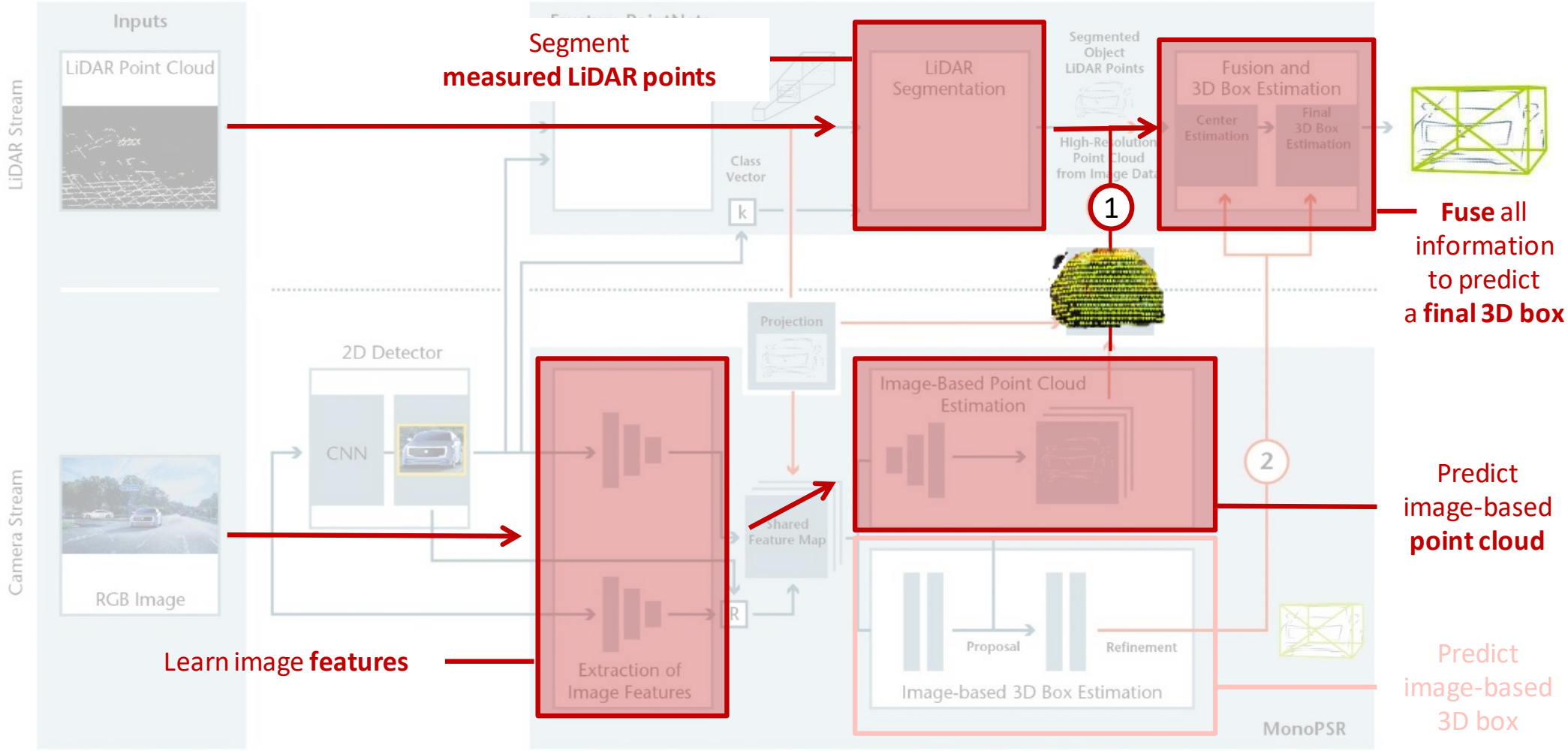
Source (Frustum PointNets): C. R. Qi et al., https://arxiv.org/abs/1711.08488
Source (MonoPSR): J. Ku et al., https://arxiv.org/abs/1904.01690

Our Approach: PSR-PointNets



Source (Frustum PointNets): C. R. Qi et al., https://arxiv.org/abs/1711.08488
Source (MonoPSR): J. Ku et al., https://arxiv.org/abs/1904.01690

PSR-PointNets – Fusion of Image-Predicted and Measured Point Clouds



Source (Frustum PointNets): C. R. Qi et al., https://arxiv.org/abs/1711.08488
Source (MonoPSR): J. Ku et al., https://arxiv.org/abs/1904.01690

PSR-PointNets – Fusion of Image-Predicted and Measured Point Clouds

- Fusion method 1: **fusion of predicted and measured point clouds**

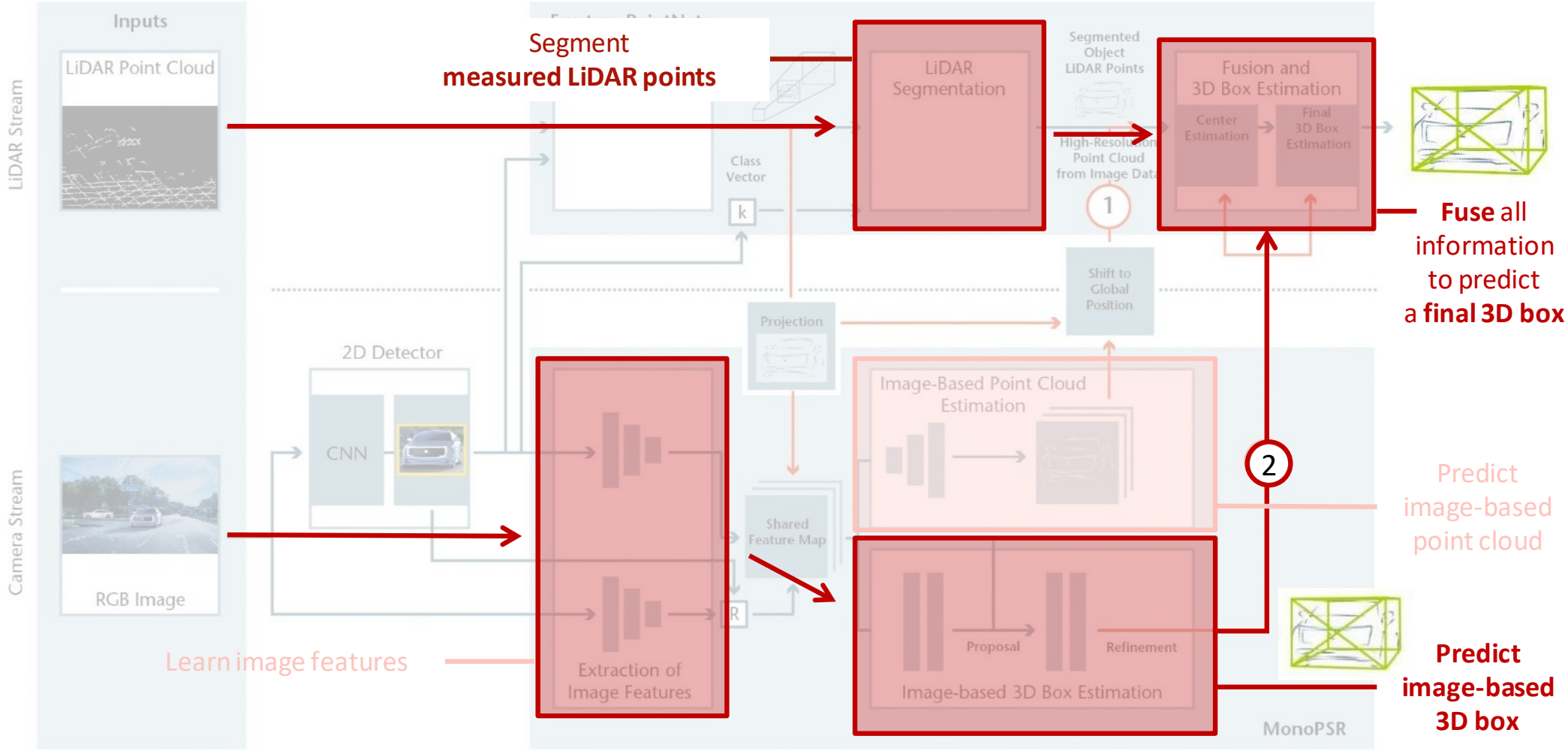
- **Significant improvement** over LiDAR-only baseline (**8 % – 10 %** absolute for moderate & hard)
- **Additional improvement** by the graph-based depth correction algorithm (GDC) from **Pseudo-LiDAR++ paper**

Detection Algorithm	Average Precision 3D (IoU = 0.7)			Average Precision 3D (IoU = 0.5)		
	Easy	Moderate	Hard	Easy	Moderate	Hard
Baseline: Frustum PointNets (LiDAR only)	50.2	38.0	32.9	86.9	67.3	59.4
MonoPSR (Image only)	10.5	10.2	8.1	46.1	35.0	31.8
PSR-PointNets – Predicted Point Cloud	45.0	41.1	37.4	87.5	75.8	69.1
PSR-PointNets – Predicted Point Cloud + GDC	**57.3**	45.7	41.2	87.1	77.0	**69.4**
PSR-PointNets – Image Center + Num. of Meas. Points	56.8	**47.6**	**42.1**	**88.5**	**77.2**	69.0

+ 8-10 %

Source (Pseudo-Lidar++): Y. You et al., Pseudo-LiDAR++: Accurate Depth for 3D Object Detection in Autonomous Driving, https://arxiv.org/abs/1906.06310, 2019

PSR-PointNets – Fusion of the Image Center



Source (Frustum PointNets): C. R. Qi et al., https://arxiv.org/abs/1711.08488
Source (MonoPSR): J. Ku et al., https://arxiv.org/abs/1904.01690

PSR-PointNets – Fusion of the Image Center

- Fusion method 2: **fusion of the image center + the number of measured LiDAR points** into the LiDAR stream

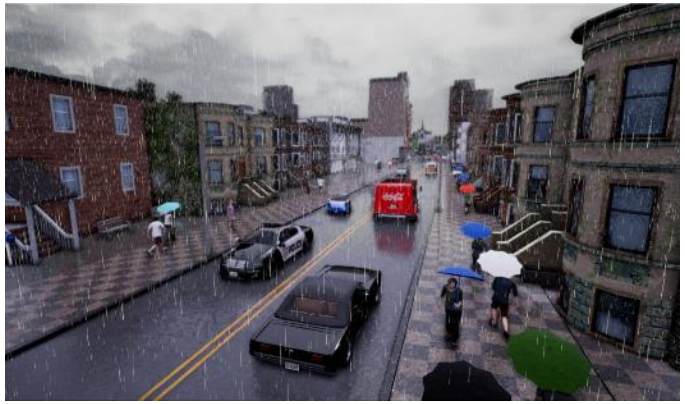
- **Significant improvement** over LiDAR-only baseline **(≈ 10 %** absolute for moderate & hard)
- Number of measured LiDAR points -> important parameter for the fusion
  -> Tells whether image or LiDAR information should be weighted higher

Detection Algorithm	Average Precision 3D (IoU = 0.7)			Average Precision 3D (IoU = 0.5)		
	Easy	Moderate	Hard	Easy	Moderate	Hard
Baseline: Frustum PointNets (LiDAR only)	50.2	38.0	32.9	86.9	67.3	59.4
MonoPSR (Image only)	10.5	10.2	8.1	46.1	35.0	31.8
PSR-PointNets – Predicted Point Cloud	45.0	41.1	37.4	87.5	75.8	69.1
PSR-PointNets – Predicted Point Cloud + GDC	**57.3**	45.7	41.2	87.1	77.0	**69.4**
PSR-PointNets – Image Center + Num. of Meas. Points	56.8	**47.6**	**42.1**	**88.5**	**77.2**	69.0

+ 10 %

Approaches to Synthetic Data

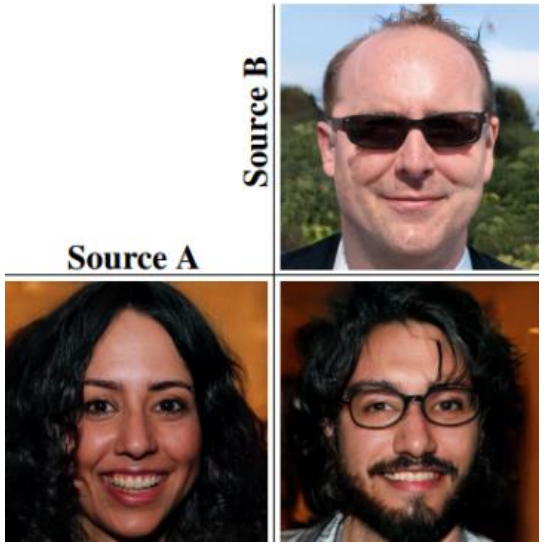
3D Engines



Source: CARLA: An Open Urban Driving Simulator, A. Dosovitskiy et. al, https://arxiv.org/pdf/1711.03938.pdf

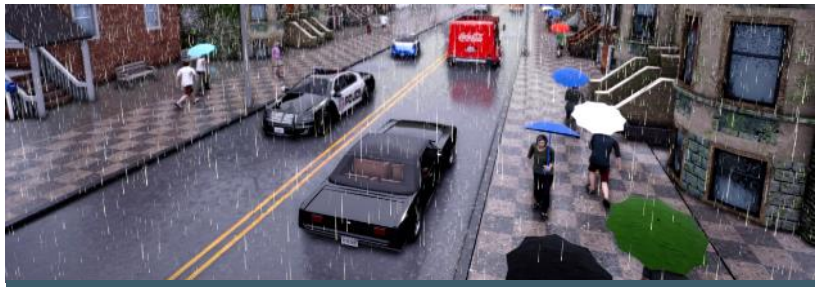
Copy & Paste Approaches

Generative Neural Networks



Source (StyleGAN): T. Karras et al., A Style-Based Generator Architecture for Generative Adversarial Networks, https://arxiv.org/abs/1812.04948

Generation of Synthetic Data – Current Trends



Often used approach: 3D engines

Drawbacks:

- **Video-game-like look**
- **Not photorealistic enough** for advanced camera-based neural networks
- **Human effort shifts**: from labeling to generating 3D worlds

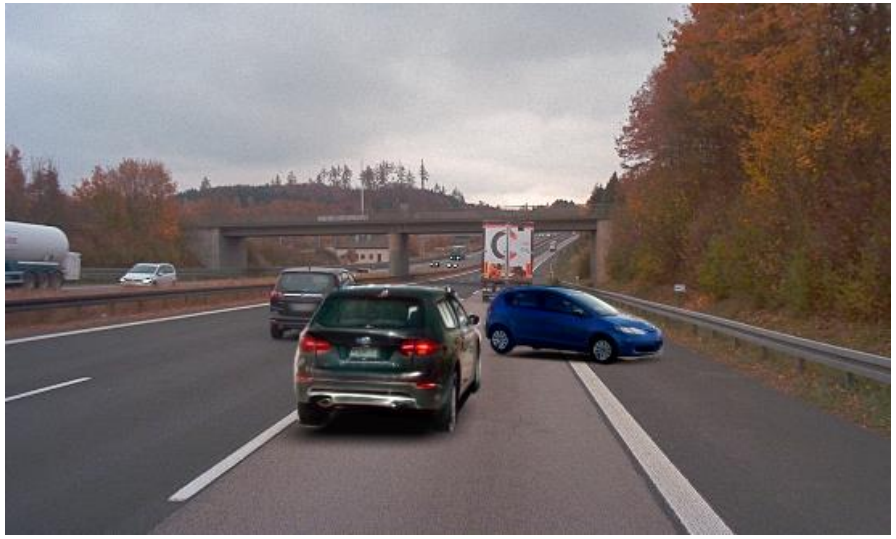
Source: CARLA: An Open Urban Driving Simulator, A. Dosovitskiy et. al, https://arxiv.org/pdf/1711.03938.pdf

Trend: Generate synthetic data based on real data

- Examples: Waymo, Baidu, Uber/Waabi
- **Copy & paste** approach for a large variety of scenarios
  -> efficient, automated, scalable with fleet data
- **Generative neural networks** for a large variety of objects
  -> very photorealistic

Source for the basic dataset: J. Geyer et al., A2D2: Audi Autonomous Driving Dataset, https://www.a2d2.audi

Our Approach To Synthetic Data – Examples



Advantages

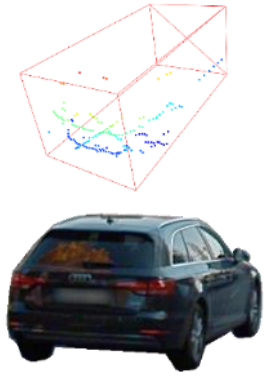
- **Consistent camera** and **LiDAR** data
- Adaptable to **every LiDAR sensor configuration**
- Based on **real data**
- **Generative adversarial neural networks** (GAN) to generate a large variety of synthetic objects
- **Automated** and **scalable** change of object properties and scenarios for **AI training** and **validation**

Source for the basic dataset: J. Geyer et al., A2D2: Audi Autonomous Driving Dataset, https://www.a2d2.audi

Generating Photorealistic Consistent Synthetic Camera And Lidar Data

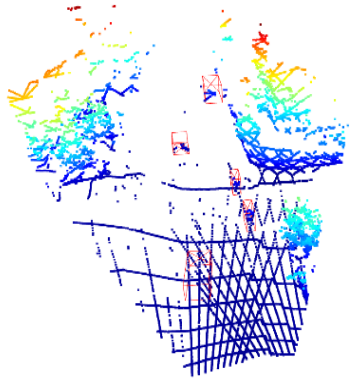
1. Annotated dataset

2. Generate foreground objects



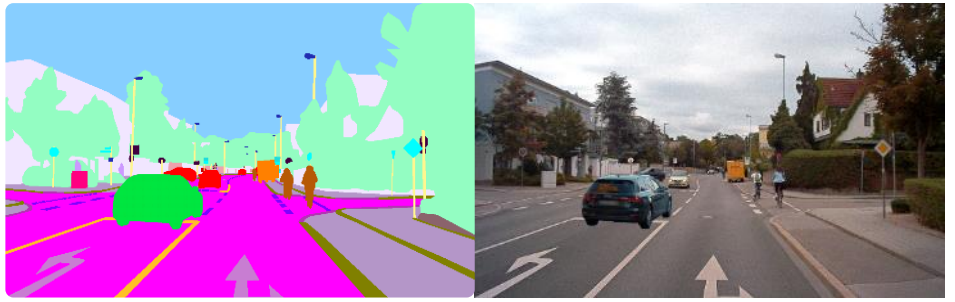
Optional: Change foreground image features with a GAN

3. Find a position in 3D space



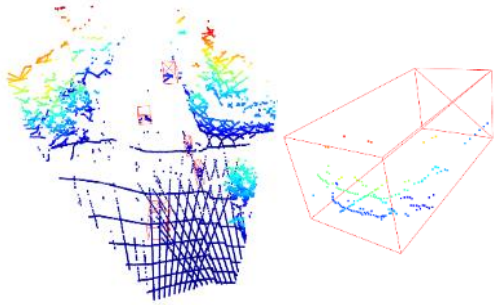
4. Project the 3D box into the image

5. Project the segmented object into the image (+ add shadow, process edge, etc.)



6. Predict a high-resolution object PC from image information (MonoPSR)

7. Adjust the predicted object PC to sensor configuration & position



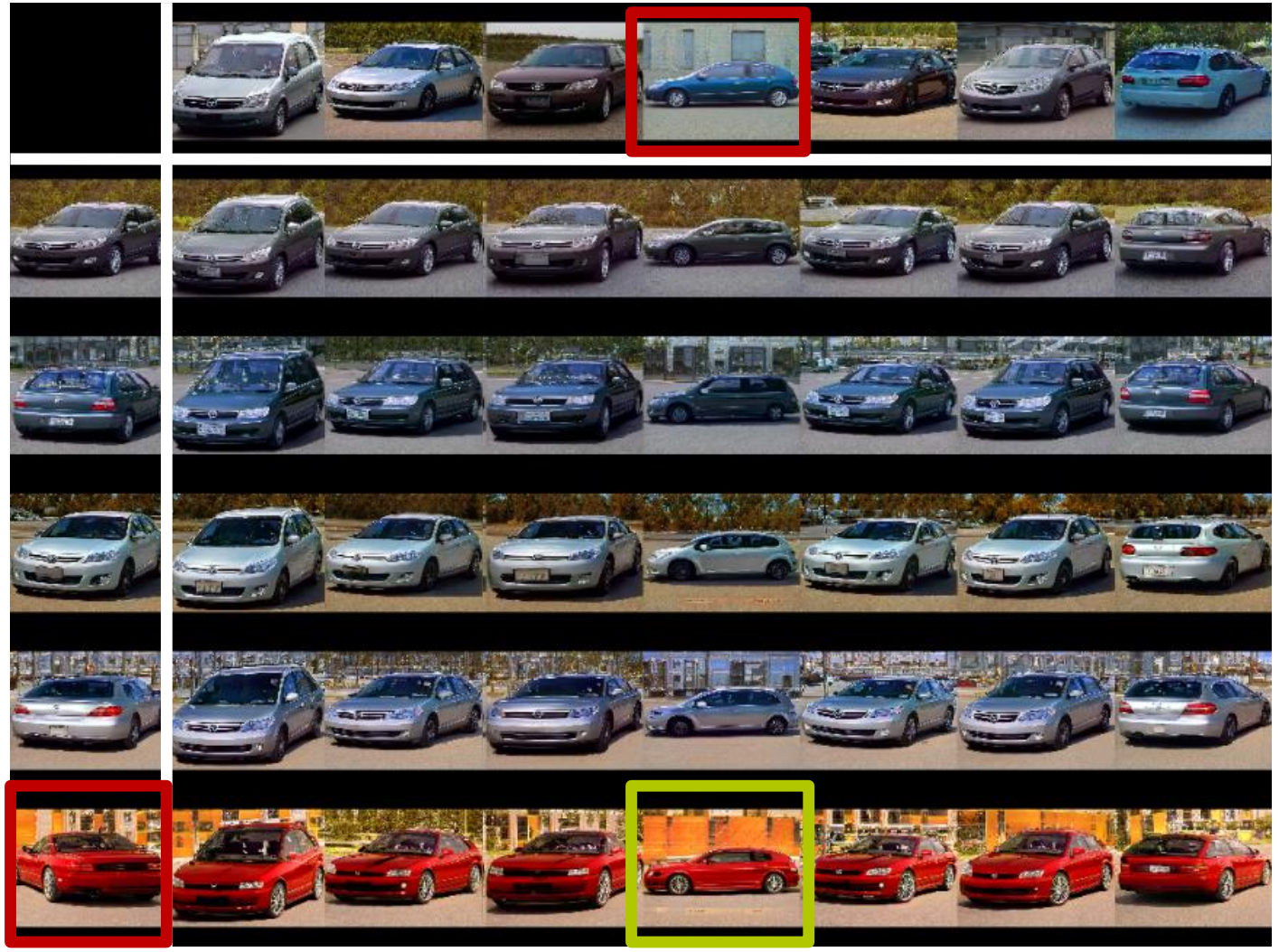
Source (A2D2 dataset): J. Geyer et al., A2D2: Audi Autonomous Driving Dataset, https://www.a2d2.audi

Generative Adversarial Networks (GANs) for Synthetic Data

Input images for …

… fine styles

… coarse styles



GAN-Generated Objects

- Generation of **synthetic objects** with **GANs** for training & testing neural networks
- Applied approaches (for research purposes):
  - StyleGAN2 (https://arxiv.org/abs/1912.04958)
  - Closed-Form Factorization (https://arxiv.org/abs/2007.06600)

Synthetic Data For 3D Object Detection – Training Results

- Algorithm: MonoPSR (**image only**)
- Scenario: **Low amount of data** available
- Baseline: Training of MonoPSR with **5% of the A2D2 dataset**
- Synthetic data: Usage of **5% of the A2D2 scenes for generation of synthetic data**
- Experiments show:
  1) Synthetic data as addition to real data can improve the performance
  2) Training only on synthetic data is equal to the performance of training with real data (in this scenario)

Dataset	Average precision (IoU = 0.5) evaluated on real data
Baseline: Real data (5% of A2D2 dataset)	27.7
Real data + synthetic data	**33.2**
Only synthetic data	28.6

Source (MonoPSR): J. Ku et al., https://arxiv.org/abs/1904.01690

Summary

- 3D object detection: Trend towards AI-based sensor fusion
- Fusion with image data improves 3D object detection with LiDAR sensors
- Synthetic data: Trend towards AI-based methods based on real data
- Efficient approach to synthetic data: Combining copy & paste methods with generative neural networks

Special thanks to
the state of Lower Saxony & Nbank
for the funding within the initiative
„Initiative Neustart Niedersachsen“

www.bertrandt.com/
ki-deep-learning

Thank you for your attention!

Are there questions?