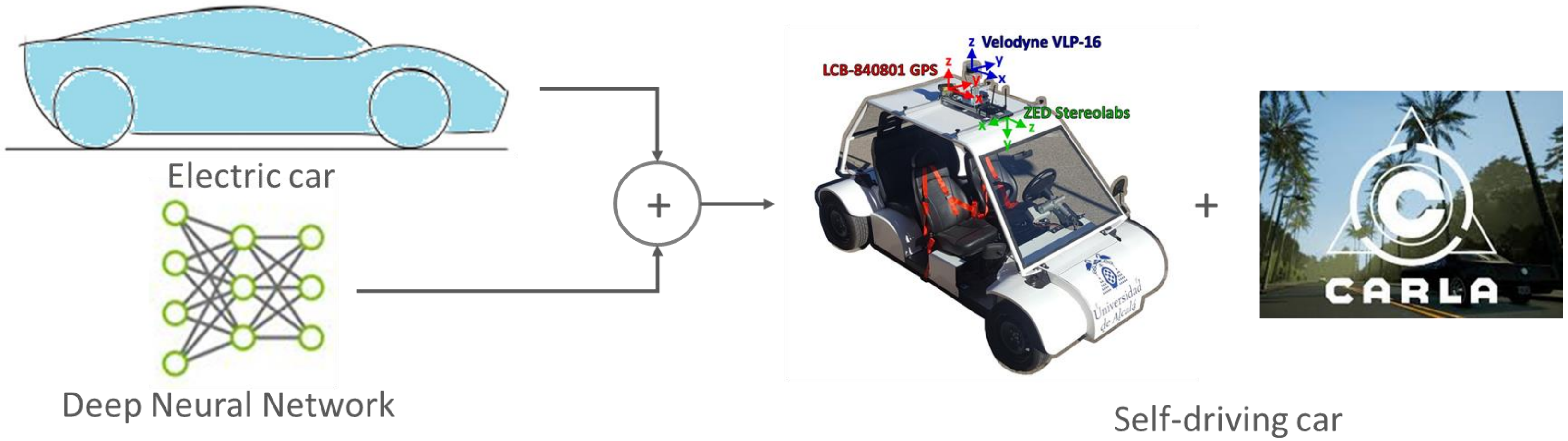
Beyond supervised Deep Learning for Autonomous Driving

Luis M. Bergasa

Robesafe Lab. University of Alcalá. Spain

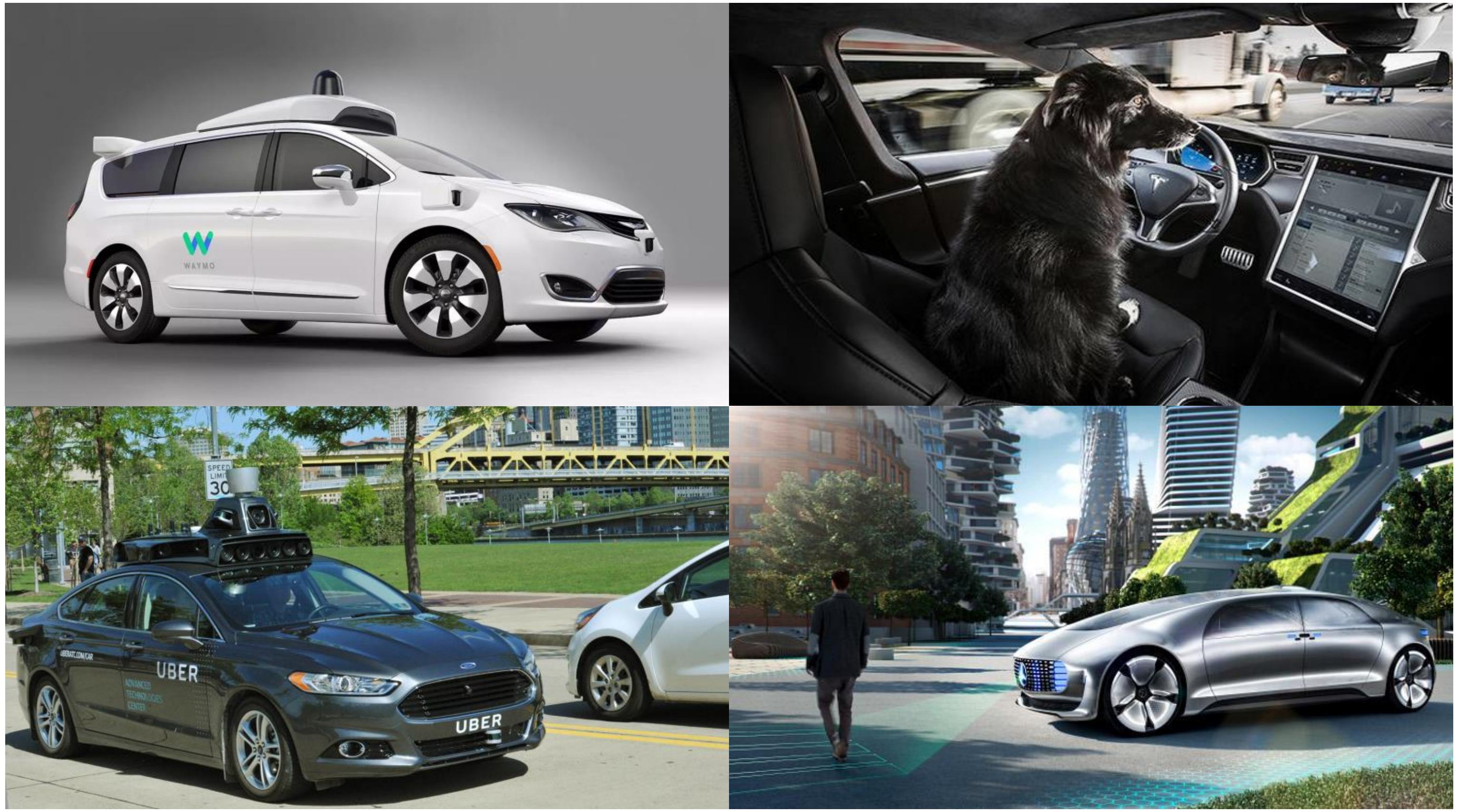


Universidad de Alcalá

RobeSafe

Introduction

Self-driving cars: Safe lives, increase mobility and improve efficiency



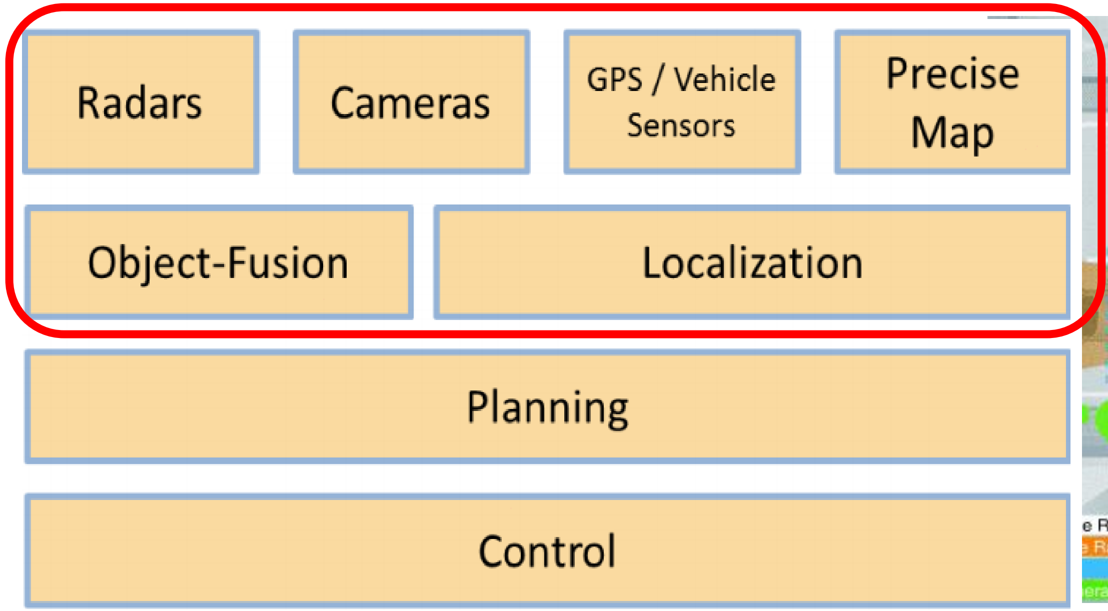
Autonomous Driving architecture

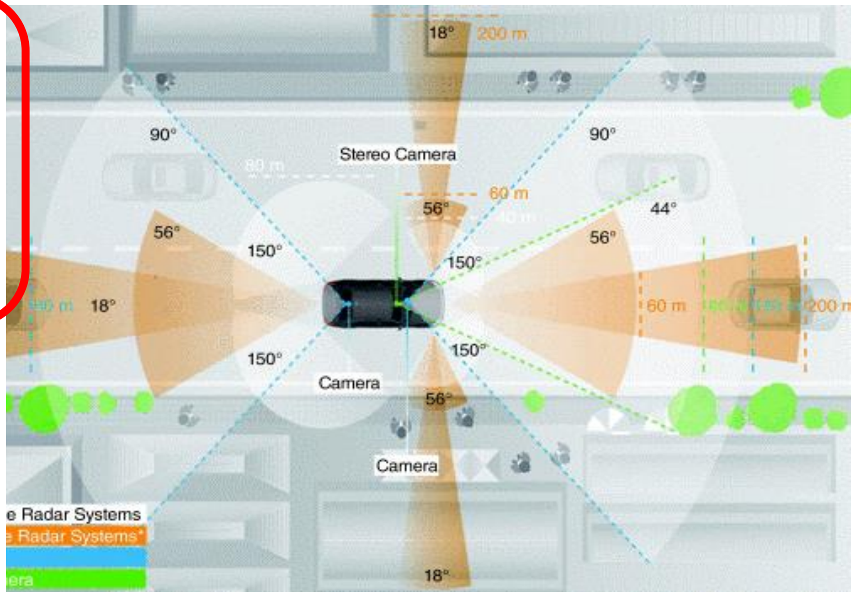
Mapping ⟶ Perception ⟶ Localization ⟶ Planning ⟶ Control

↑

Decision Making

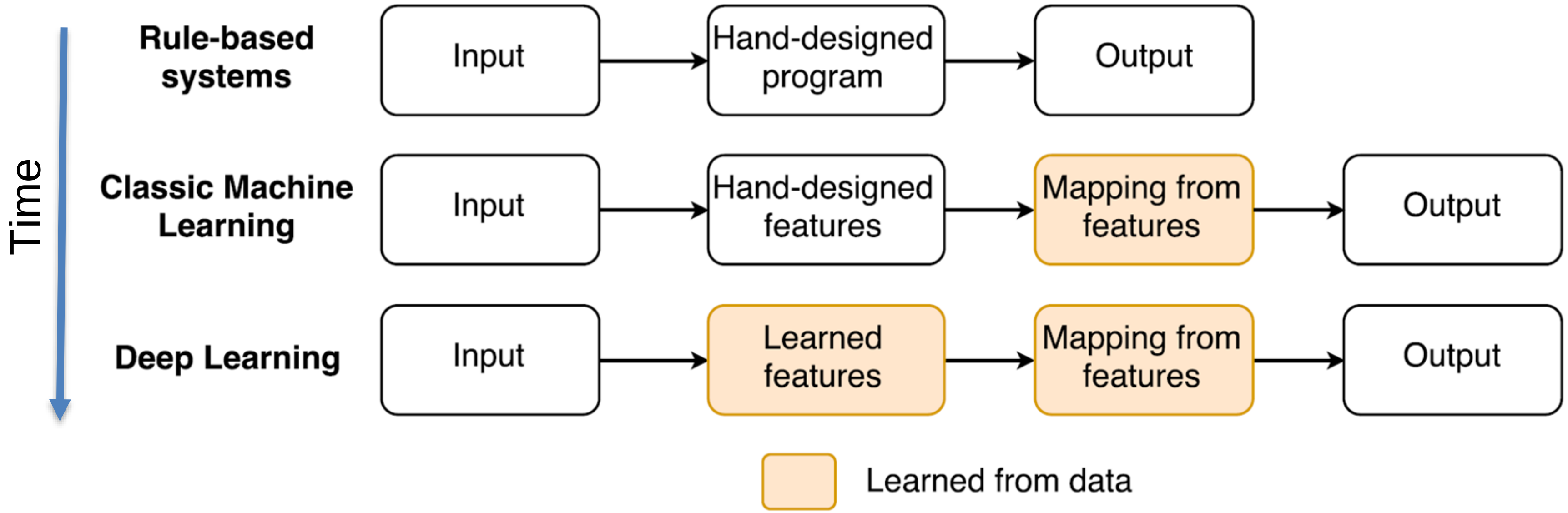
Perception





Introduction

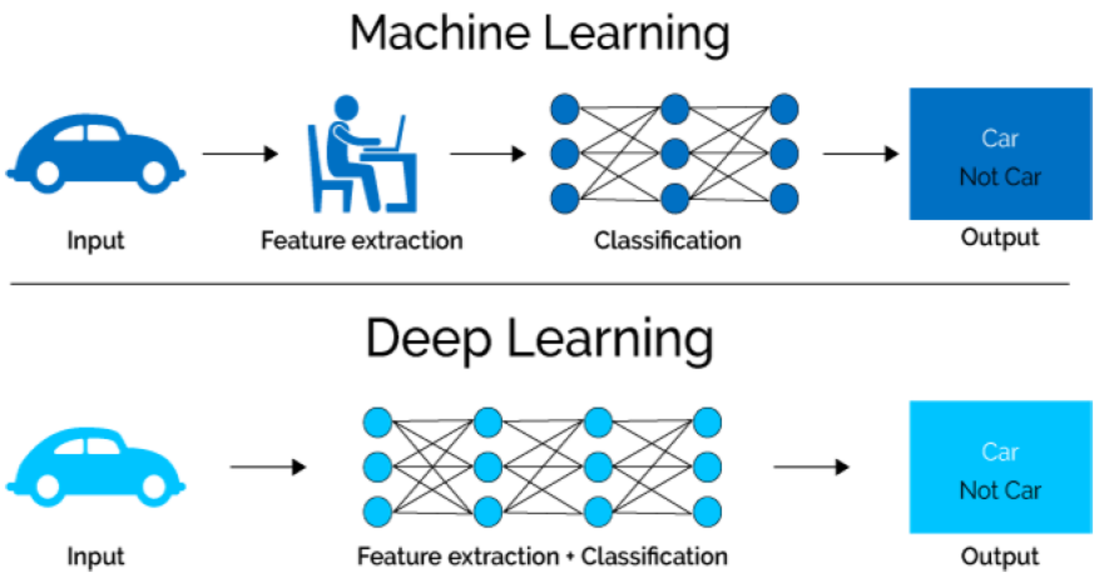
3 approaches for AD design in time:

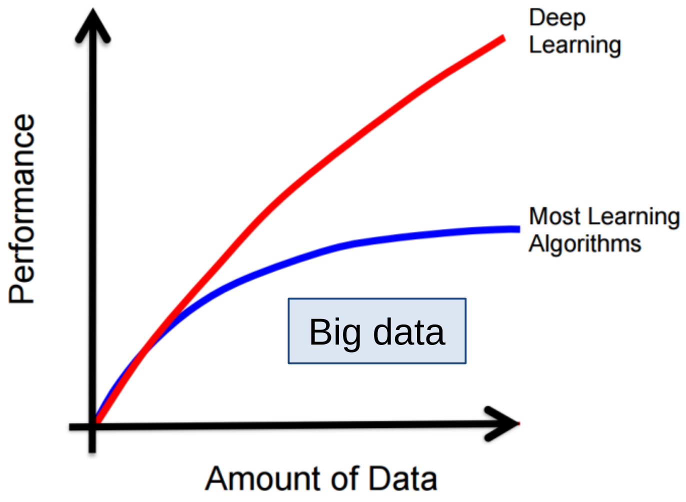


From **Model-driven** to **Data-driven** algorithms

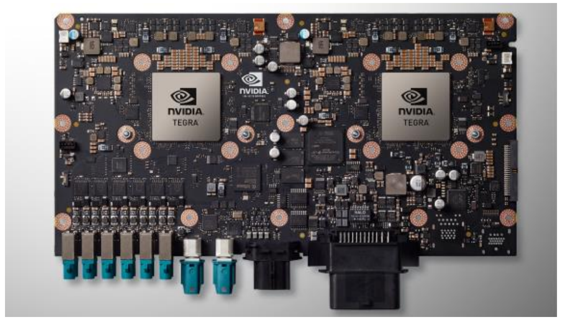
Introduction

Why Deep Learning? Scalable Machine Learning and **Parallelizable** processing



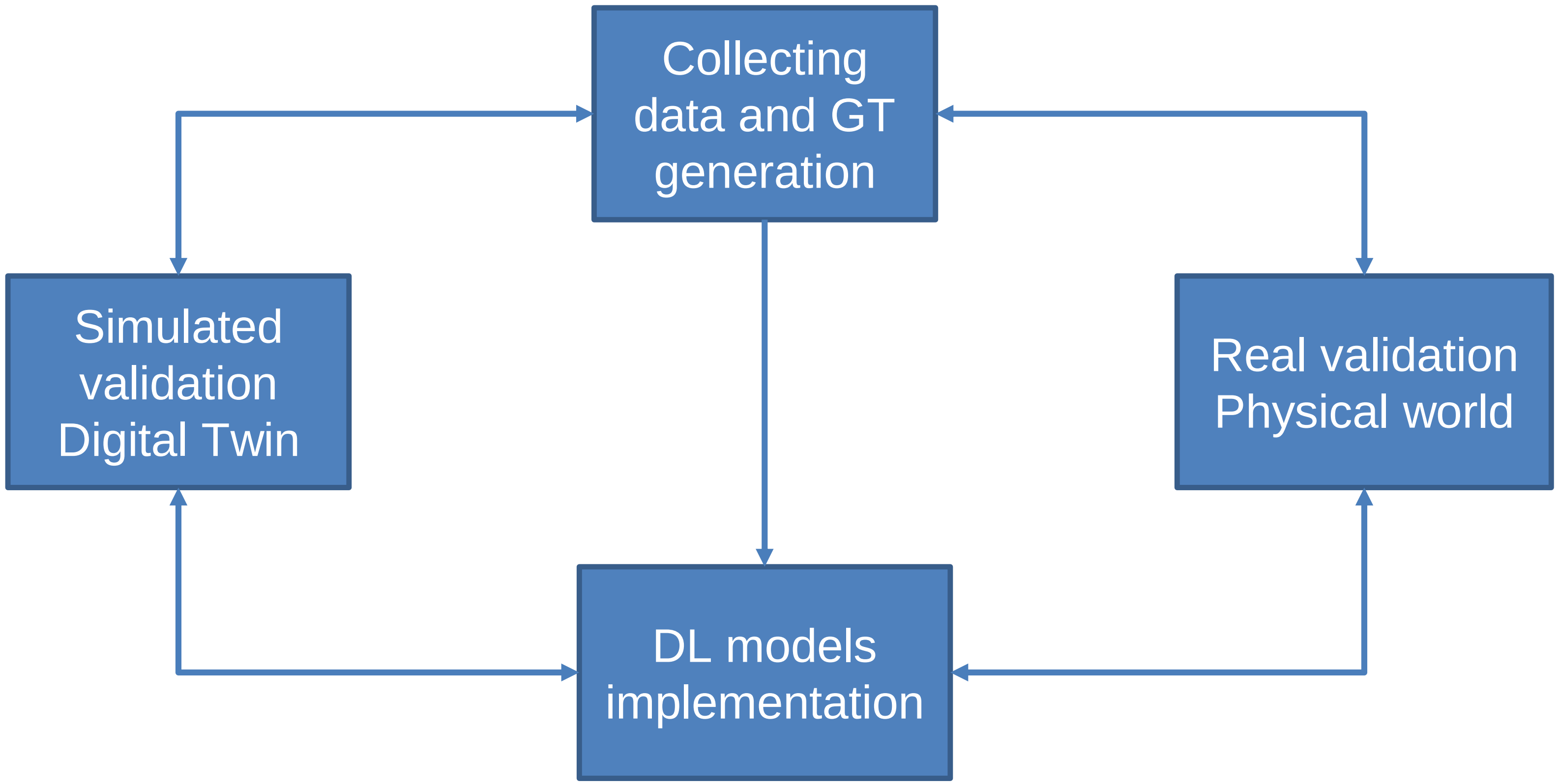


Big data



Specific HW
CPU -> GPU -> TPU

AD stack implementation pipeline in the context of DL



Supervised learning paradigm

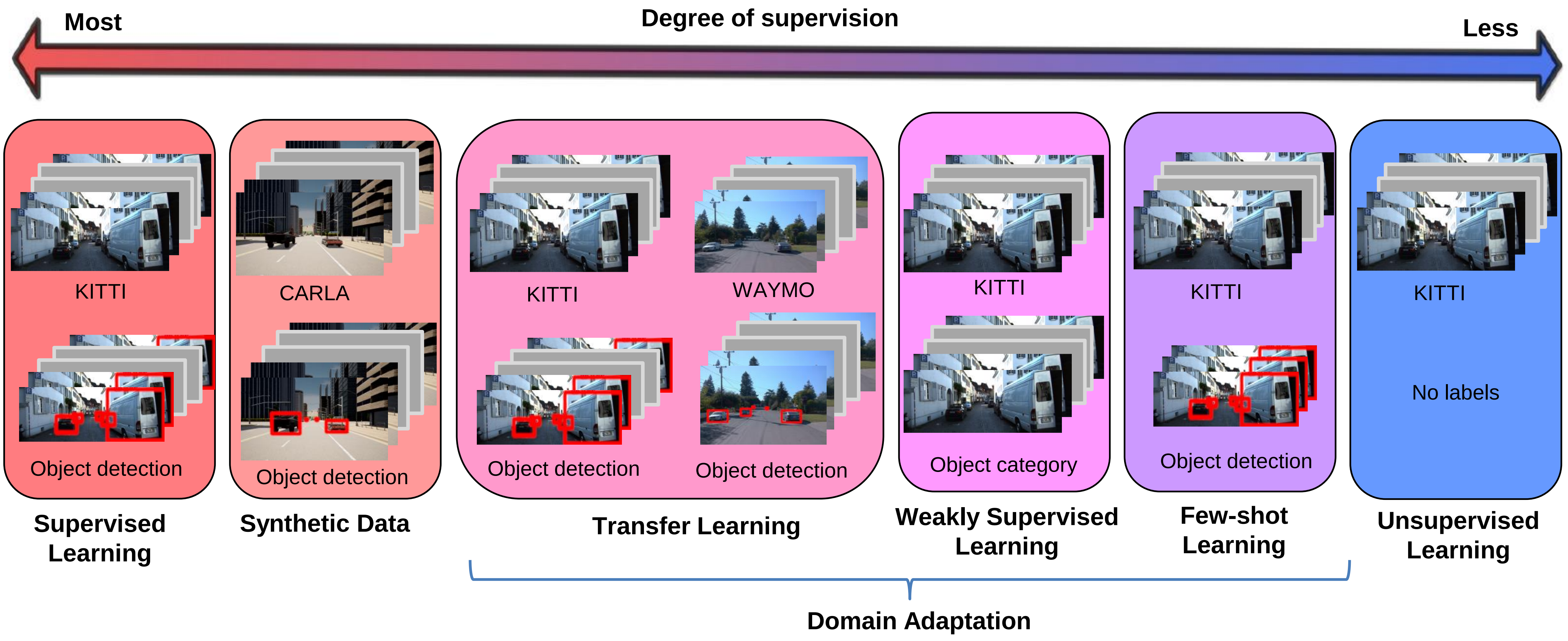
Beyond supervised DL

- **Supervised DL** have led to the development of **accurate and efficient models** for AD tasks

- Supervised DL-based techniques for AD present important **issues**:
  - Require huge amounts of human-labeled data
  - Are unable to generalize to multiple domains and tasks

- There are a plethora of works to **adapt models to label-scarce target dataset** through unsupervised domain adaptation

Beyond supervised DL



Synthetic data

- **Simulators** can be used to construct **synthetic data sets** for AD
- **Advantages**:
  1. Exponentially larger than real-world data sets
  2. Precise GT available for free and without human errors
  3. Viewpoint, lighting and material properties can be easily configured
  4. Challenging scenarios can be introduced in the dataset
     - Unsafety situations (e.g. head-on collisions)
     - Different weather conditions (e.g. sunny, rainy, foggy, …)
     - Rare situations (e.g. unexpected pedestrian)

Synthetic data

Table 4. Comparison of autonomous driving simulators

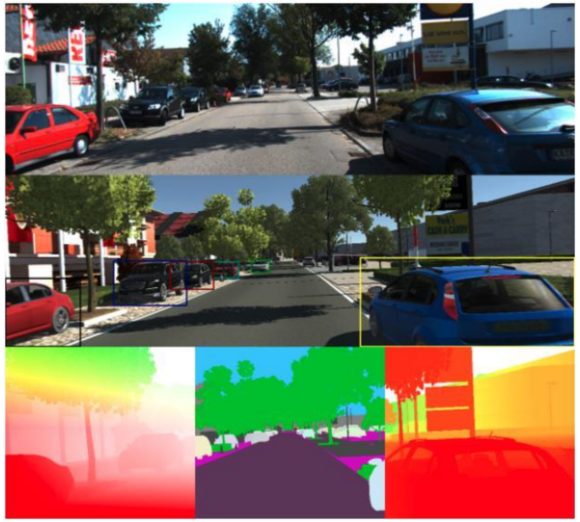
Features		CARLA	AirSim	DeepDrive	LGSVL	NVIDIA Drive	rFpro
General	Licence	Open Source	Open Source	Open Source	Open Source	Commercial	Commercial
	Portability	Windows & Linux	Windows & Linux	Windows & Linux	Windows & Linux	Windows & Linux	Windows & Linux
	Physics Engine	Unreal Engine	Unreal Engine & Unity	Unreal Engine	Unity	Unreal Engine	U
	Scripting Languages	Python	C++, Python, Java	C++, Python	Python	Python	U
Environment	Urban Driving	Town	Town, City	Road Track	City	City, Harbor	Town, City, Road Track
	Off- Road	N	Forest, Mountain	N	N	N	N
	Actors - Human	Y	N	N	Y	U	Y
	Actors - Cars	Y	Y	Y	Y	U	Y
	Weather Conditions	Y	Y		Y	Y	Y
Sensors	RGB	Y	Y	Y	Y	Y	Y
	Depth	Y	Y	Y	Y	U	Y
	Thermal	N	Y	N	N	U	U
	LiDAR	Y	Y	N	Y	Y	Y
	RADAR	Y	N	N	Y	Y	Y
Output Training Labels	Semantic Segmentation	Y	Y	N	Y	Y	Y
	2D Bounding Box	Y	N	N	Y	Y	U
	3D Bounding Box	Y	N	Y	Y	U	U

Legend: U: Unknown, Y: Yes, N: No

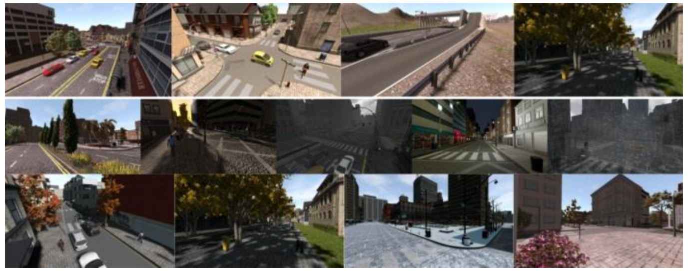
[Malik et al, 2022] Malik, Sumbal, Manzoor Ahmed Khan, and Hesham El-Sayed. "CARLA: Car Learning to Act—An Inside Out." *Procedia Computer Science* 198 (2022): 742-749.

Synthetic data

- Virtual KITTI is a photo-realistic synthetic video dataset designed to learn and evaluate computer vision models for several video understanding tasks
- SHYNTIA is a synthetic dataset that consists of 9400 multi-viewpoint photo-realistic frames rendered from a virtual city and interesting for semantic segmentation studies
- CARLA is an open-source simulator for autonomous driving research



Source: https://arxiv.org/pdf/1605.06457.pdf.

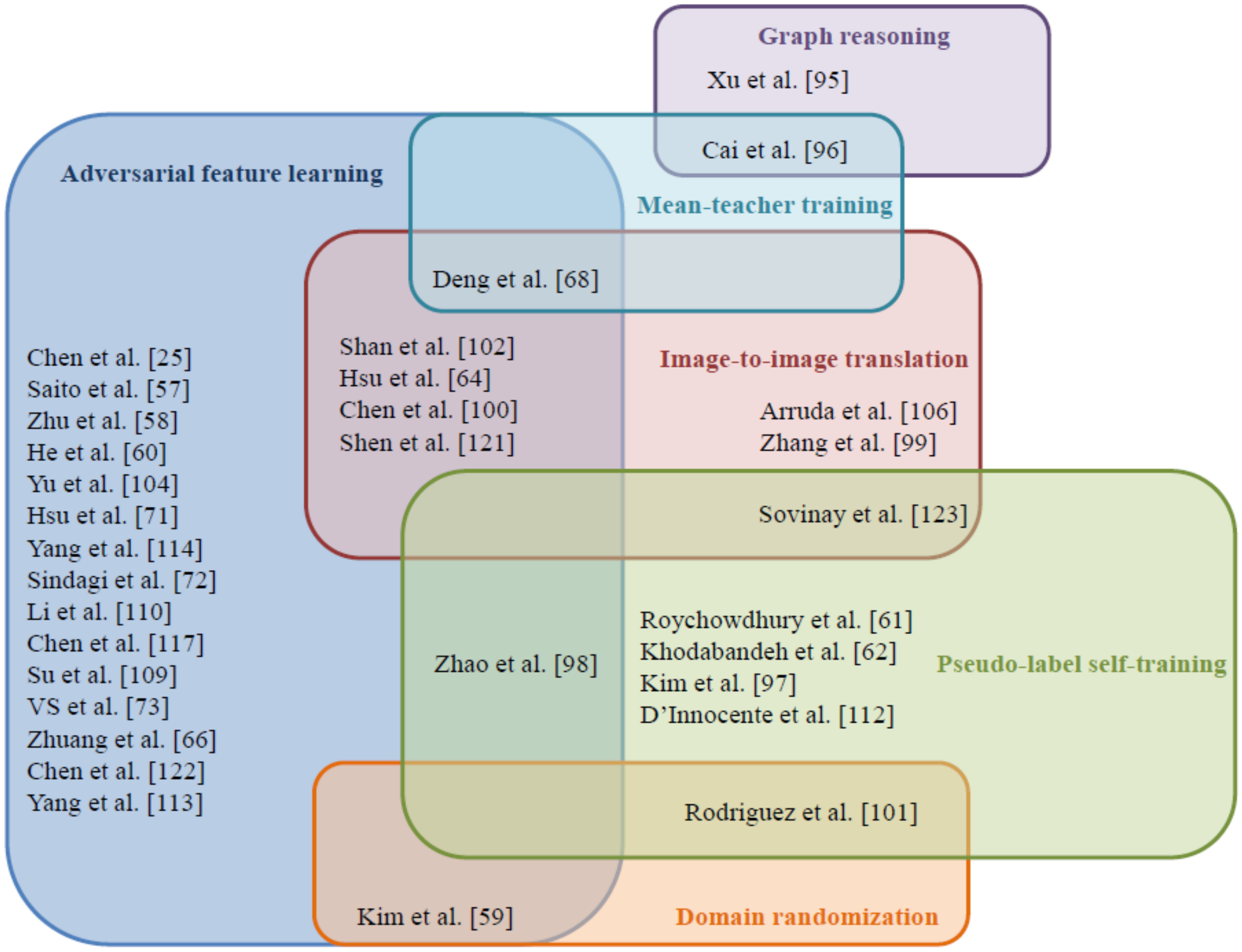


Source: https://synthia-dataset.net/.

Source: https://synthia-dataset.net/.

Challenge: domain gap between real and synthetic data sets

Domain adaptation



[Oza et al, 2021] "Unsupervised domain adaptation of object detectors: A survey", Oza P, Sindagi VA, VS V, Patel VM. arXiv preprint arXiv:2105.13502. 2021 May 27.

Domain adaptation research directions

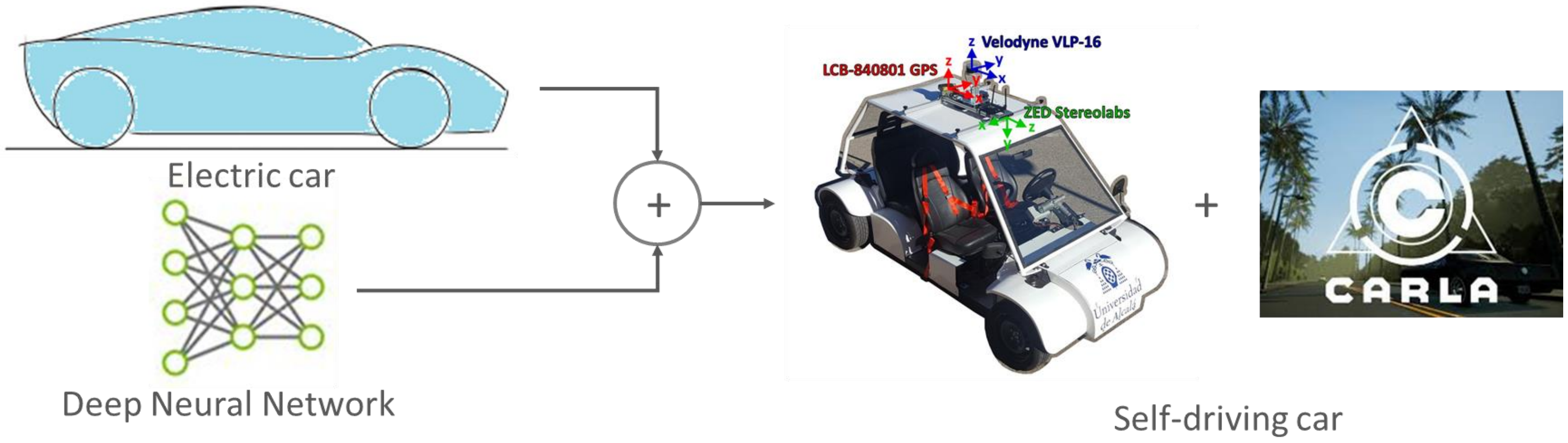
- **Comprehensive evaluation**
  - Generalizing current approaches to **other detection frameworks** (YOLO, 3D detectors, etc.)
  - Design **complex real-world datasets** (to build datasets reflecting real-world scenarios that enable a more rigorous and robust evaluation process)
  - **Consistent training and inference strategy** (hyper-parameters related to the backbone net and resizing ratio of the input images) to perform a fair comparison of the methods

Domain adaptation research directions

- **Improving generalization with real-world constraints**
  - **Weak/semi-supervised domain adaptation**
    - It is important to explore further to bridge the performance gap between fully supervised and adaptive training
  - **Few-shot domain adaptation**
    - Is not yet explored in the domain adaptive detection literature
  - **Multi-source domain adaptation**
    - Training data are collected from multiple sources. Selective adaptation can be carried out where only relevant samples are considered
  - **Continuous time adaptation**
    - Adapting a pretrained model to dynamically changing environmental conditions

Our approach

To contribute as university researchers **to apply DL techniques to special AVs in towns**



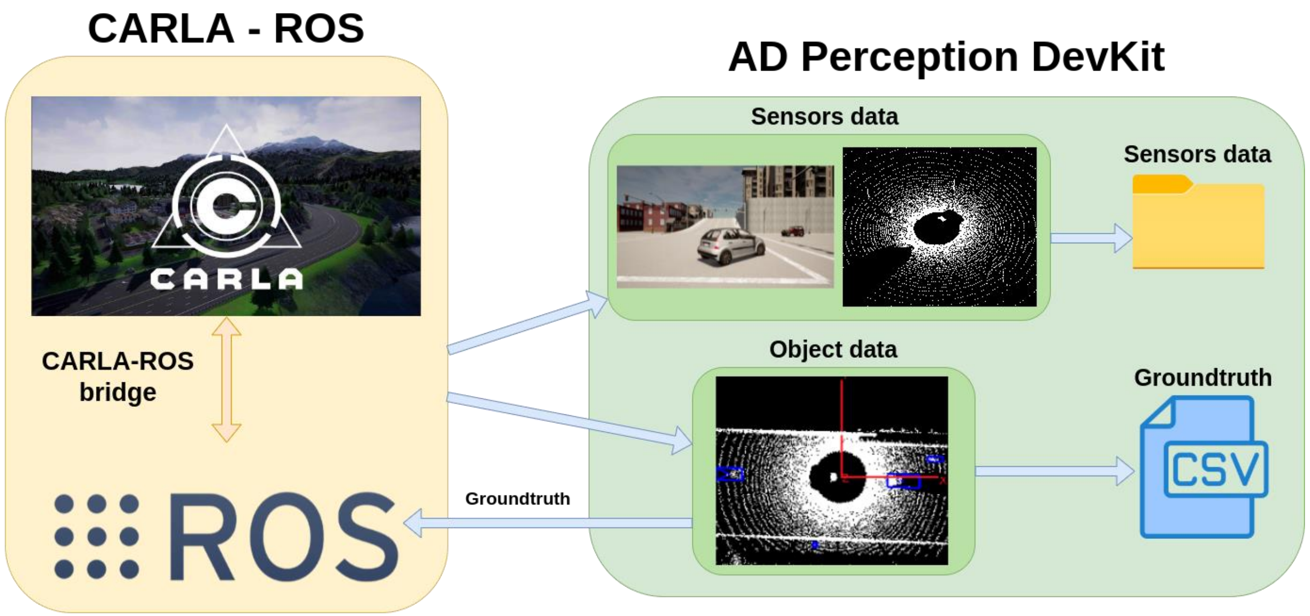
- **Goals**:
  - Take advantage of the open-source car chassis, datasets, frameworks, DL models and simulators to build a real AV
  - Implement a classic AD stack and progressively introduce DL in its modules
  - Validate our contributions in simulation and in our real car

Our contributions beyond SDL

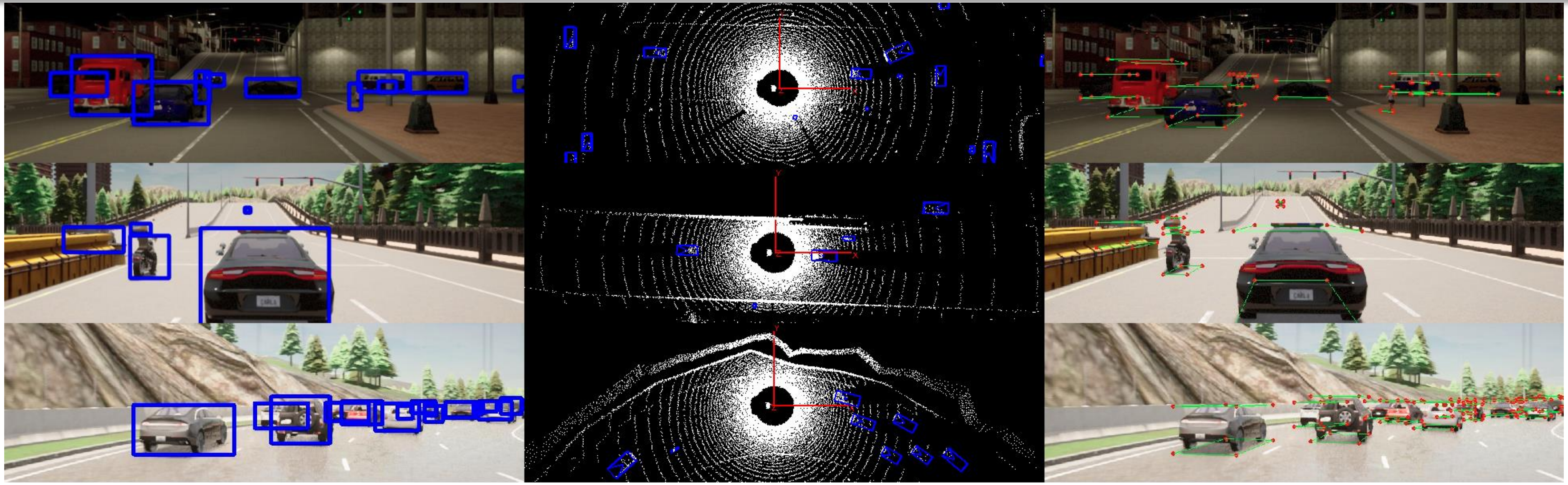
Contributions / Research directions	1 AD Perception Development Kit	2 DA for 2D object detection	3 DA for 3D object detection	4 DA for SS	5 Autonomous Driving Stack validated in CARLA Leaderboard
Generalizing to other detection frameworks		X	X	X	X
Complex real-world datasets	X				X
Fair comparison	X	X	X	X	X
Few-shot domain adaptation		X	X	X	
Multi-source domain adaptation		X	X	X	
Continuous time adaptation					X

1. AD Perception Development Kit

- Based on **CARLA** simulator
- Able to generate an **infinite set of annotated data** referred to the ego-vehicle
- **Recording virtual car sensors** (camera, LiDAR, Radar) in **virtual environments**
- **Dataset** generated with this tool for some specific **challenging scenarios**



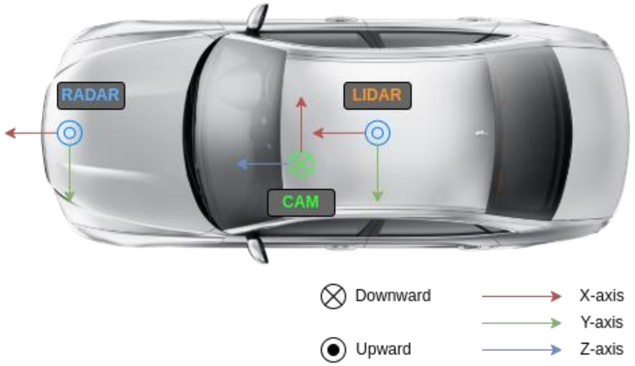
1. AD Perception Development Kit



2D BEV 3D

TABLE IV: Sensor data in the AD PerDevKit dataset.

Sensor	Brief Details
Camera	Front stereo camera at 20Hz, with a FoV of 85º and a resolution of 1280x720 generating RGB images.
LiDAR	360º of visibility, a maximum range of 120m, 64 beams at 20Hz and a vertical FoV of 2º to -24.9º, generating 1,300,000 points per second.
Radar	Frontal radar at 20Hz with a horizontal FoV of 90º, vertical FoV of 18º and a maximum range of 150m, generating 9,000 points per second.



1. AD Perception Development Kit

TABLE I: Comparison of cited datasets.

Dataset	Tipe of data	Cities	Annotated frames	Ground-truth 360º	Camera	LiDAR	Radar	Night/Rain	**GT generation tool**
KITTI	Real data	1	15k	✗	✓	✓	✗	✗	✗
nuScenes	Real data	2	40k	✓	✓	✓	✓	✓	✗
AIODrive	Synthetic data	8	100k	✓	✓	✓	✓	✓	✗
AD DevKit	Synthetic data	5	21k	✓	✓	✓	✓	✓	✓

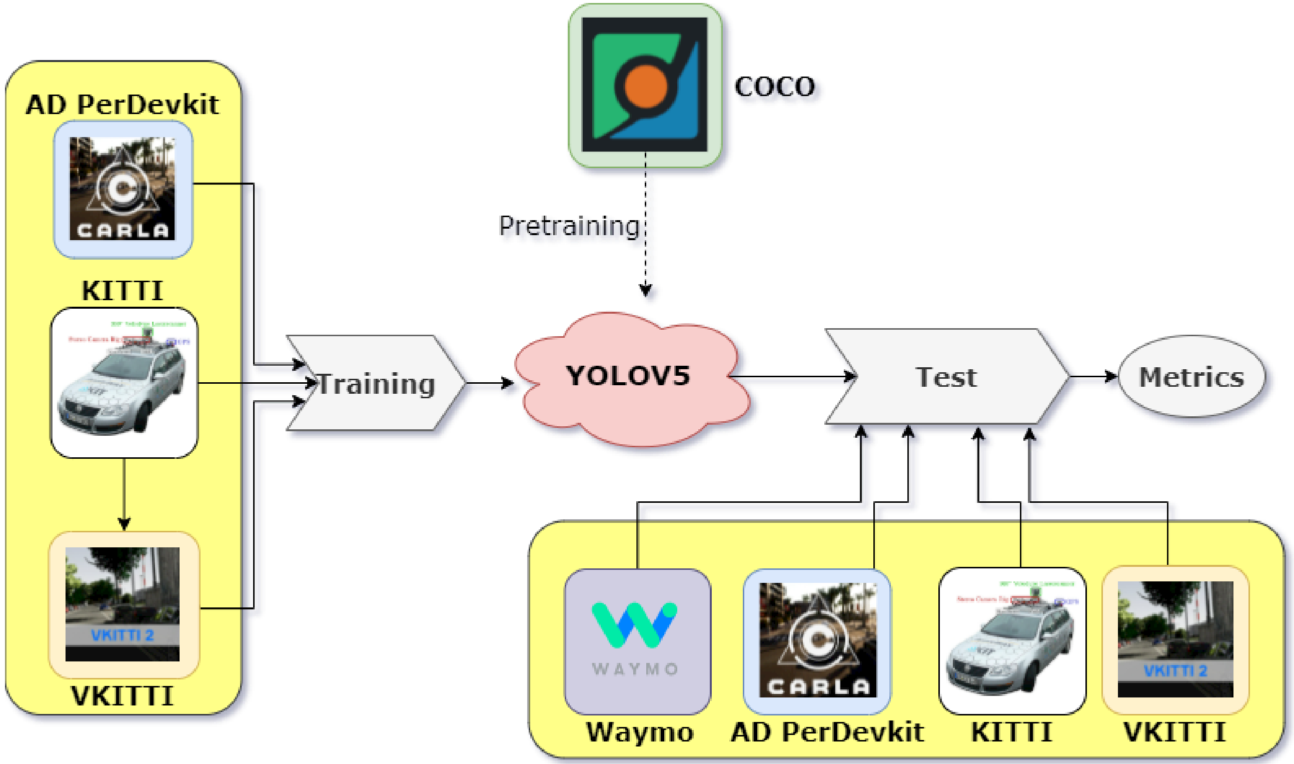
TABLE III: Dataset information

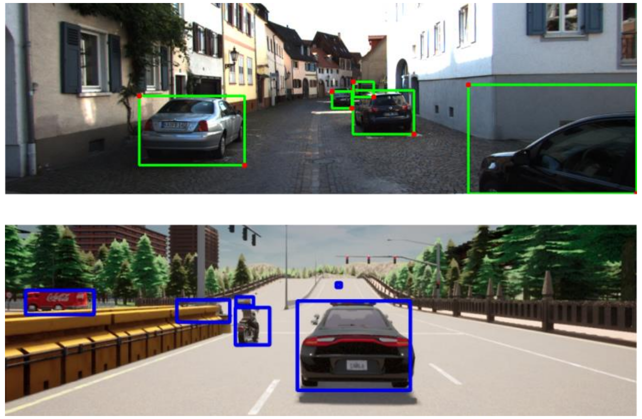
Town	Weather	Challenging scenarios	Frames	Objects per frames
03	Day	Intersection and roundabout	1708	29.11
	Night		1822	45.89
05	Day	Crowded intersections	3249	19.53
		Highway	2078	9.58
06	Day	Entrance to highway	1413	8.21
		Crowded highway	1673	16.1
	Rain		1487	15.86
07	Day	Small village	2738	10.89
	Rain		2560	9.78
10HD	Day	Crowded town	1587	77.9
	Night		1409	76.81

[Peña et al, 2022] "AD PerDevKit: An Autonomous Driving Perception Development Kit using CARLA simulator and ROS", J. de la Peña, L. M. Bergasa, M. Antunes, F. Arango, C. Gómez-Huélamo, and E. LópezGuillén, **IEEE Conference on Intelligent Transportation Systems (ITSC) 2022,** Macau, China, October 2022. In revision
GitHub repository: https://github.com/Javier-DlaP/ad_perdevkit

2. Domain adaptation for a 2D object detector

- Study **domain adaptation** in a **2D image detector trained with real world and synthetic data** from multiple sources





KITTI

CARLA

[Antunes et al, 2022] "Including transfer learning and synthetic data in a training process of a 2D object detector for autonomous driving", M. Antunes, L. M. Bergasa, J. Araluce, R. Gutiérrez, F. Arango and M. Ocaña, **IEEE Conference on Intelligent Transportation Systems (ITSC) 2022,** Macau, China, October 2022. In revision

2. Domain adaptation for a 2D object detector

- **Experiment 1:**
  - Detector trained on one particular dataset (source dataset) does not generalize well to a dataset that has a different distribution (target dataset)
  - Applying a pre-training backbone and fine tuning the heads with the target domain obtains higher performance
  - Synthetic data generalize worse over real images than using few-shot learning with 10% of target dataset

Model	Train	Test	P	R	mAP
Yolov5L	Kitti	Kitti	0.924	0.869	**0.943**
		Waymo	0.618	0.350	0.409
		Carla	0.276	0.198	0.154
	Carla	Kitti	0.078	0.077	0.038
		Waymo	0.023	0.023	0.013
		Carla	0.621	0.893	0.780
	VKitti	Kitti	0.681	0.416	0.462
		Waymo	0.048	0.047	0.017
		VKitti	0.982	0.941	0.984

From scratch

Model	Train	Test	P	R	mAP
Yolov5L	Kitti 100%	Kitti	0.953	0.910	**0.969**
		Waymo	0.776	0.582	0.674
		Carla	0.409	0.559	0.439
	Kitti 50%	Kitti	0.921	0.867	0.942
		Waymo	0.786	0.568	0.693
	Kitti 25%	Kitti	0.869	0.781	0.883
		Waymo	0.798	0.622	0.716
	Kitti 10%	Kitti	0.821	0.708	0.802
		Waymo	0.793	0.643	0.737
	Carla	Kitti	0.651	0.466	0.532
		Waymo	0.614	0.509	0.552
		Carla	0.710	0.842	0.793
	VKitti	Kitti	0.797	0.622	0.656
		Waymo	0.838	0.663	0.769
		VKitti	0.966	0.935	0.983

Pretrained with COCO

2. Domain adaptation for a 2D object detector

- **Experiment 2:**
  - With pre-trained weights a mix with synthetic data does not achieve better results than training with a single real-world dataset (100%)
  - A mix of synthetic data and 10% of target real world dataset generalize better that using only the 10% of real-world target dataset

Model	Train	Test	P	R	mAP
Yolov5L	Kitti 100%	Kitti	0.953	0.910	**0.969**
		Waymo	0.776	0.582	0.674
		Carla	0.409	0.559	0.439
	Kitti 50%	Kitti	0.921	0.867	0.942
		Waymo	0.786	0.568	0.693
	Kitti 25%	Kitti	0.869	0.781	**0.883**
		Waymo	0.798	0.622	0.716
	Kitti 10%	Kitti	0.821	0.708	0.802
		Waymo	0.793	0.643	0.737
	Carla	Kitti	0.651	0.466	0.532
		Waymo	0.614	0.509	0.552
		Carla	0.710	0.842	0.793
	VKitti	Kitti	0.797	0.622	0.656
		Waymo	0.838	0.663	0.769
		VKitti	0.966	0.935	0.983

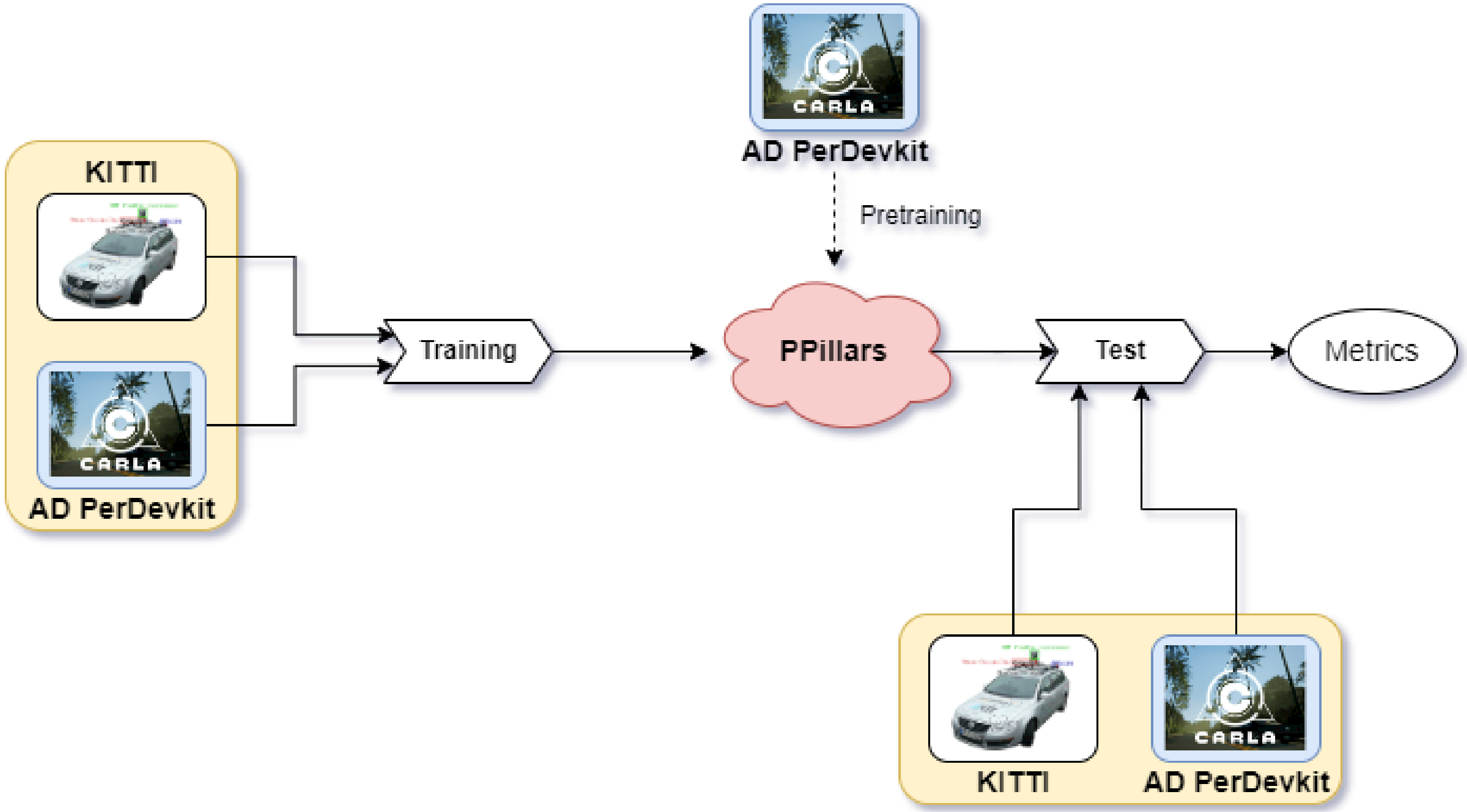
Pretrained with COCO

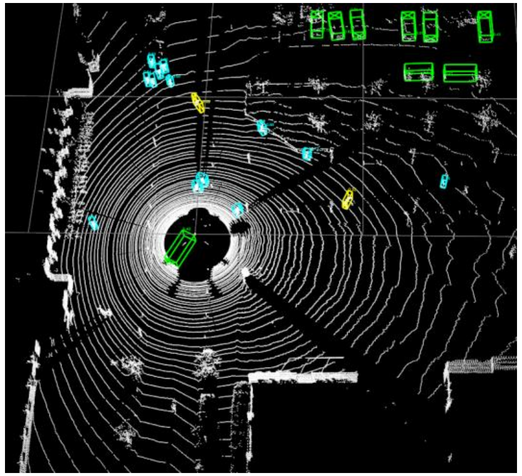
Model	Train	Test	P	R	mAP
Yolov5L	Kitti 100%+ Carla	Kitti	0.950	0.911	**0.969**
		Waymo	0.800	0.579	0.676
		Carla	0.745	0.760	0.829
	Kitti 50%+ Carla	Kitti	0.802	0.871	**0.889**
		Waymo	0.792	0.600	0.694
		Carla	0.610	0.901	0.800
	Kitti 25%+ Carla	Kitti	0.817	0.808	**0.846**
		Waymo	0.799	0.570	0.673
		Carla	0.702	0.731	0.776
	Kitti 10%+ Carla	Kitti	0.787	0.806	**0.821**
		Waymo	0.740	0.559	0.643
		Carla	0.644	0.813	0.760

Model	Train	Test	P	R	mAP
Yolov5L	Kitti 100%+ Vkitti	Kitti	0.949	0.915	**0.967**
		Waymo	0.770	0.597	0.683
		Vkitti	0.973	0.929	0.983
	Kitti 50%+ Vkitti	Kitti	0.938	0.882	**0.955**
		Waymo	0.779	0.597	0.682
		Vkitti	0.971	0.925	0.982
	Kitti 25%+ Vkitti	Kitti	0.895	0.878	**0.931**
		Waymo	0.781	0.586	0.679
		Vkitti	0.979	0.930	0.984
	Kitti 10%+ Vkitti	Kitti	0.877	0.815	**0.886**
		Waymo	0.797	0.582	0.686
		Vkitti	0.964	0.939	0.983

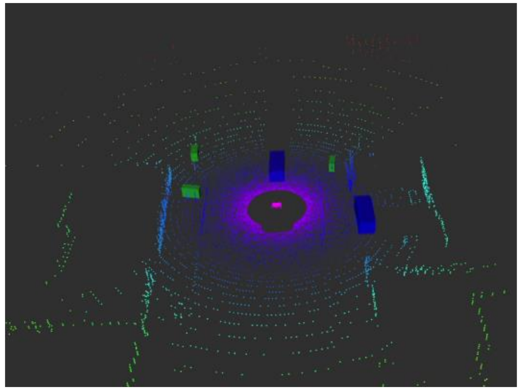
3. Domain adaptation for a 3D object detector

- Study **domain adaptation** in a **3D point cloud detector trained with real world and synthetic** data from multiple sources





KITTI



CARLA

3. Domain adaptation for a 3D object detector

- **Experiment 1**: 3D detector is capable of learning on both real and artificial data quite similarly but validation on the opposite type of data set leads to unacceptable performance
- **Experiment 2**: Combining real and synthetic data into one data set shows no significant improvement in the real world
- **Experiment 3**: Fine-tunning a model trained with synthetic data with a 20% of the target dataset improves real-world performance

Experiment 1: Train on one database

Network	Train set	Test set	Prec	Recall	F1	mAP
PPillars	Kitti	Kitti	0.89	0.80	0.85	0.74
		Carla	0.00	0.00	0.00	0.00
	Carla	Kitti	0.48	0.17	0.25	0.14
		Carla	0.96	0.66	0.78	0.62

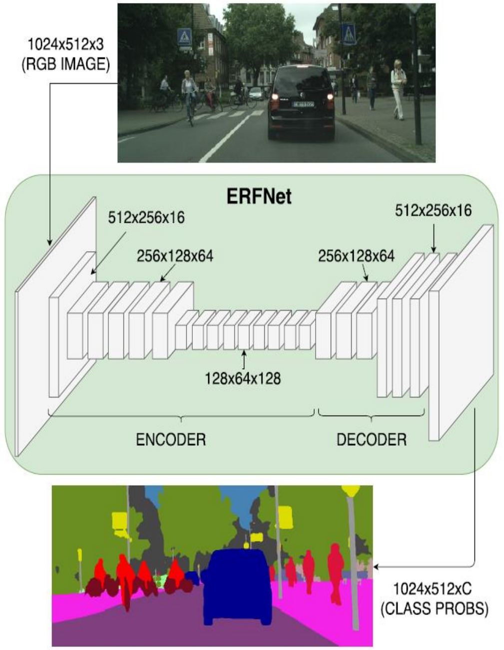
Experiment 2: Train on combined databases

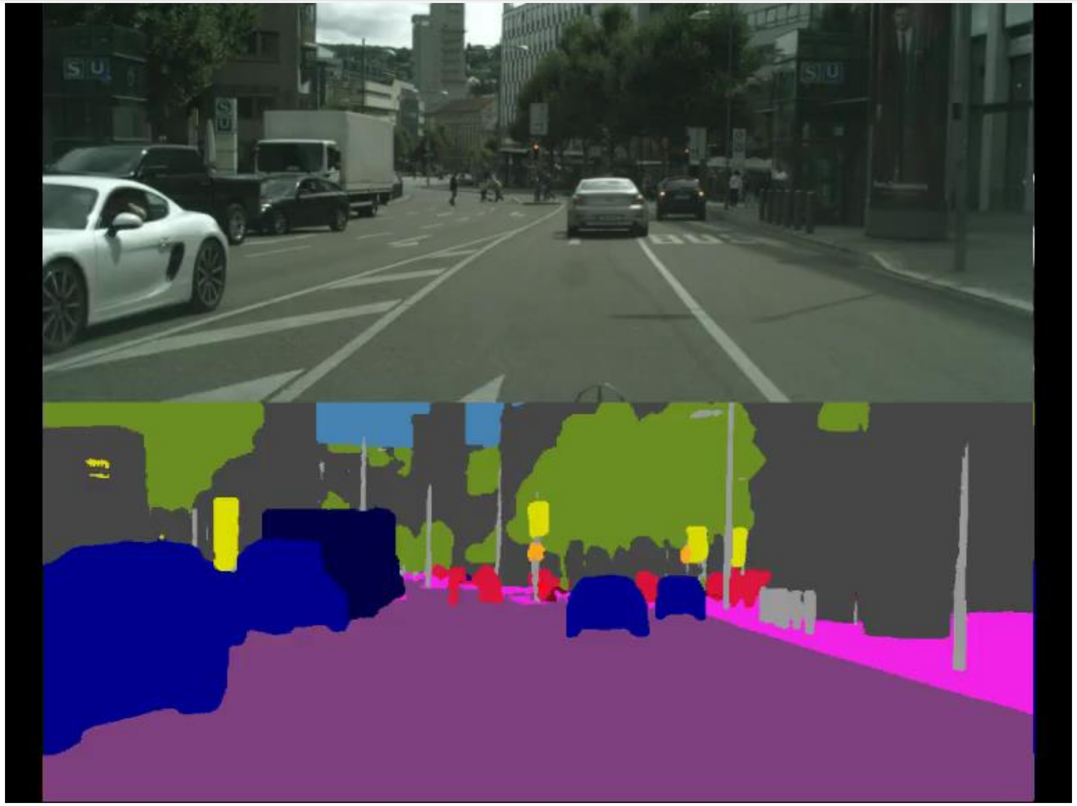
Network	Train set	Test set	Prec	Recall	F1	mAP
PPillars	Kitti 100% + Carla 100%	Kitti	0.92	0.76	0.83	0.68
		Carla	0.97	0.64	0.77	0.62
	Kitti 100% + Carla 20%	Kitti	0.94	0.74	0.83	0.69
		Carla	0.94	0.53	0.67	0.52

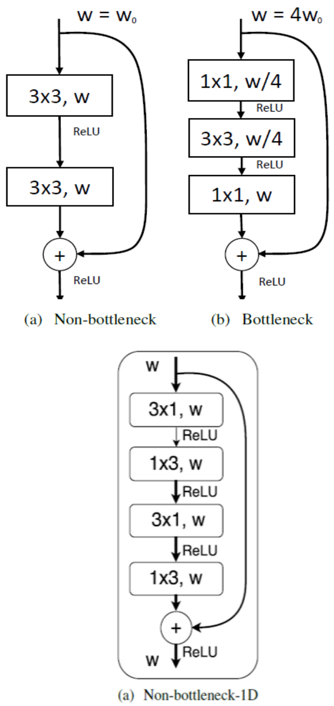
Experiment 3: Test on CARLA, fine-tune & test on KITTI

Network	Tune set	Epoch	Prec	Recall	F1	mAP
PPillars	Kitti 20%	1	0.91	0.68	0.78	0.59
		5	0.90	0.73	0.80	0.66
		10	0.85	0.79	0.82	0.63
		15	0.83	0.81	0.82	0.70
		25	0.94	0.70	0.80	0.69
	Kitti 100%	1	0.91	0.70	0.79	0.67
		5	0.87	0.79	0.83	0.65

4. Domain adaptation for our SS ERFNet





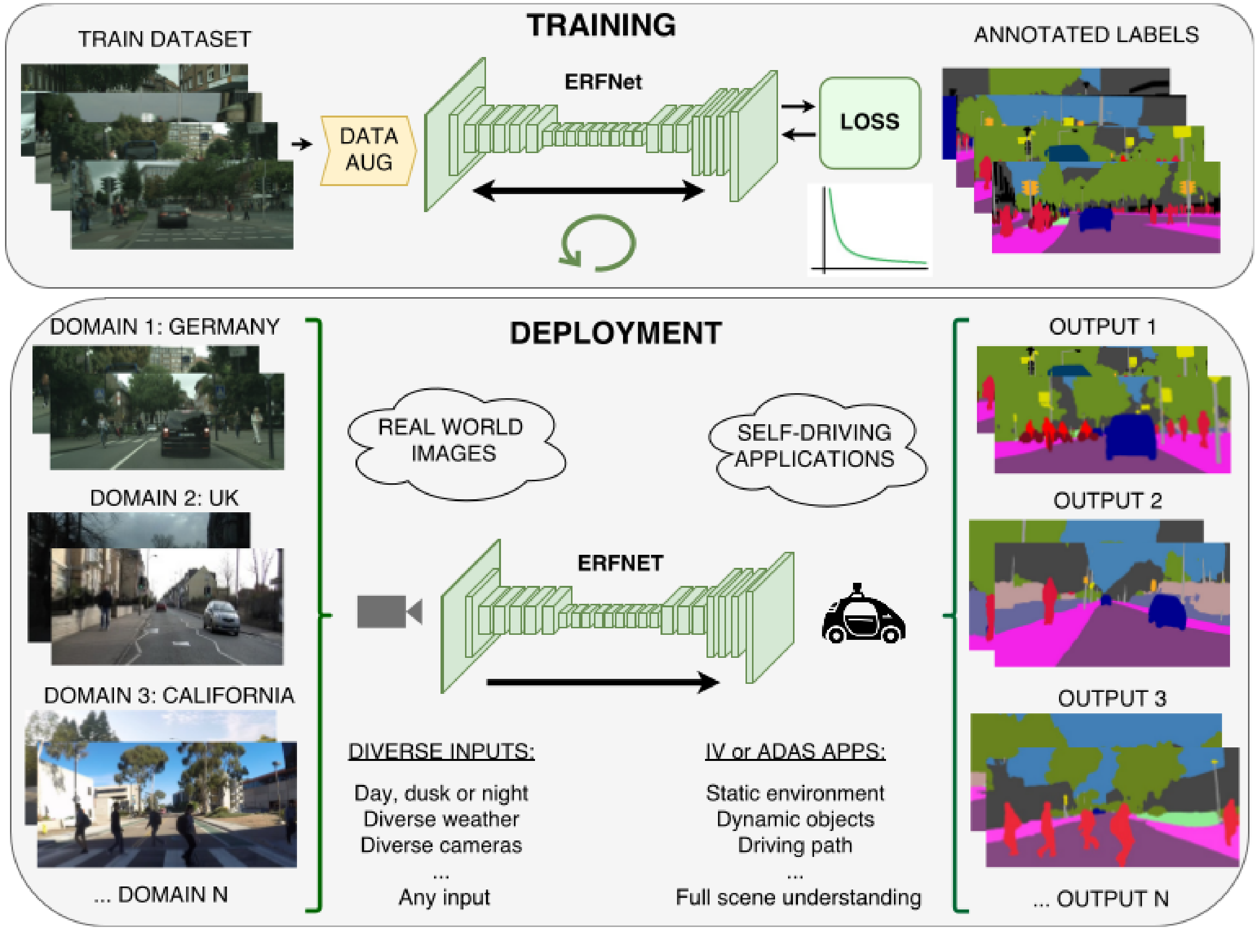


- Residual layers
- Filter factorization (1D kernels)

[Romera et al, 2018a] "ERFNet: *Efficient Residual Factorized ConvNet for Real-time Semantic Segmentation",* E. Romera, J. M. Álvarez, L. M. Bergasa and R. Arroyo, **IEEE Transactions on Intelligent Transportation Systems (T-ITS), January 2018.** **[IEEE T ITS Main publication (GSM)]**

GitHub repository: https://github.com/Eromera/erfnet

4. Domain adaptation for our SS ERFNet

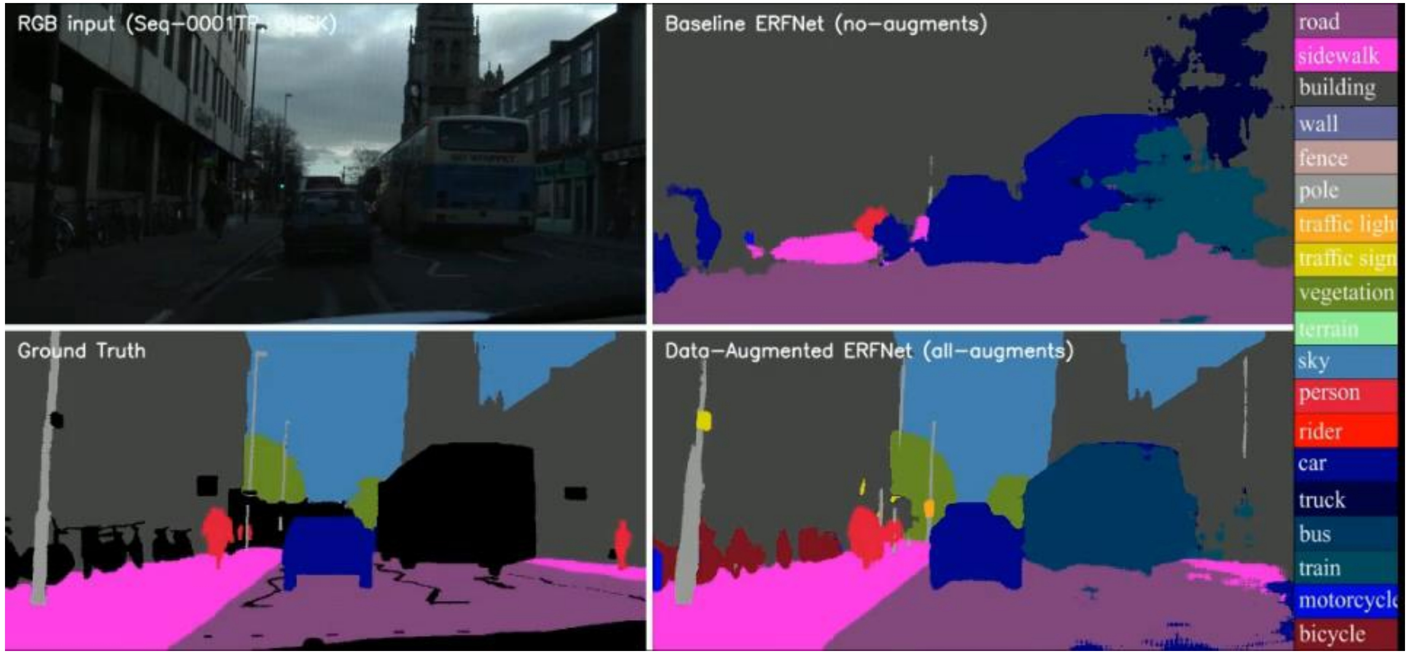


[Romera et al, 2018b] “*Train Here, Deploy There: Robust Segmentation in Unseen Domains”,* E. Romera, L. M. Bergasa, J. M. Álvarez and M. Trivedi, IEEE Intelligent Vehicles Symposium (**IV**), Changshu, Suzhou, China, June 2018.

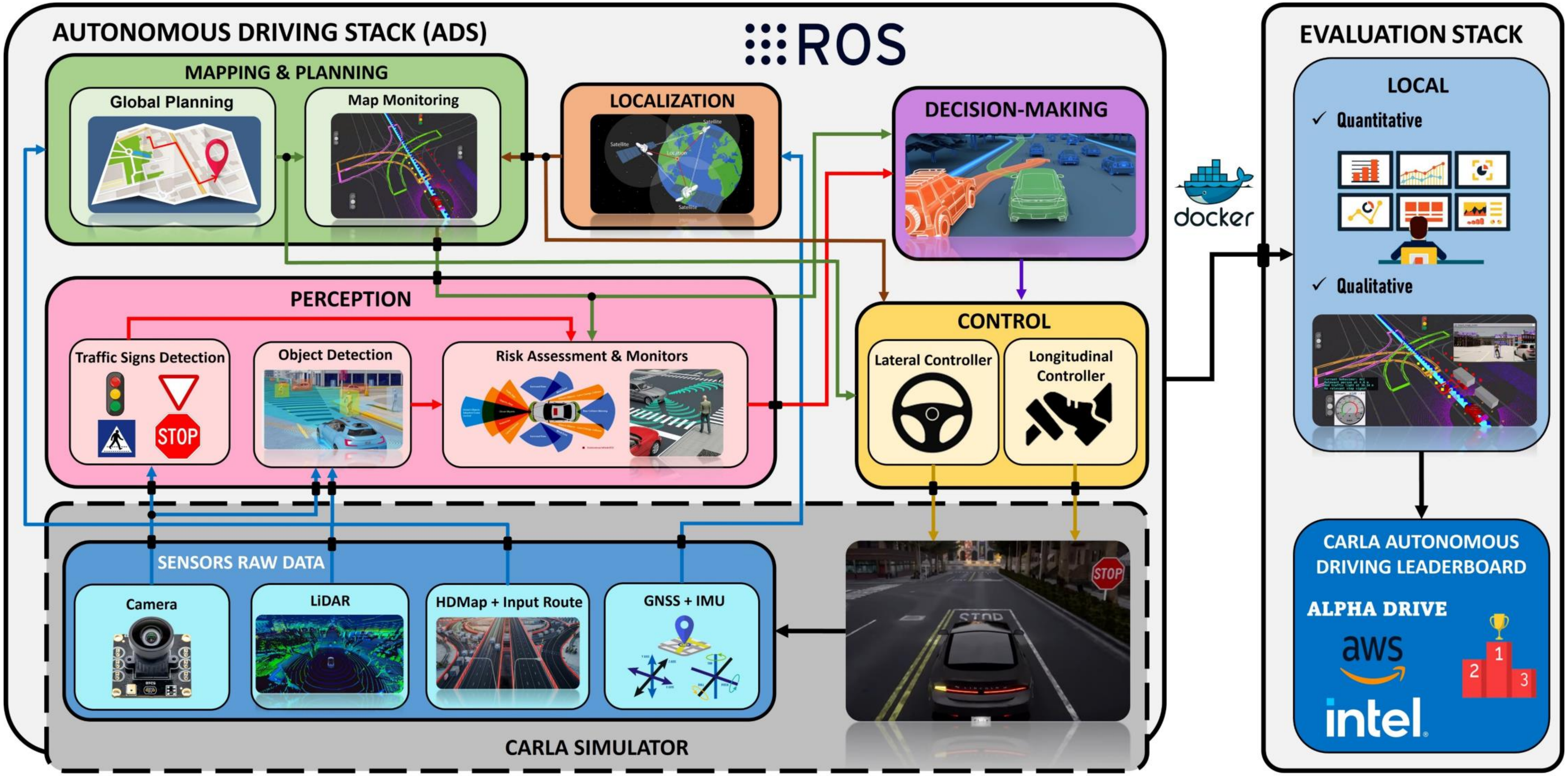
4. Domain adaptation for our SS ERFNet

3 Options to improve SS performance:

1. **Increase data samples** → Very time-expensive due to per-pixel annotations
2. **Use synthetic data** → SS model overfits the synthetic noise
3. **Perform data augmentation** → Simple techniques commonly used

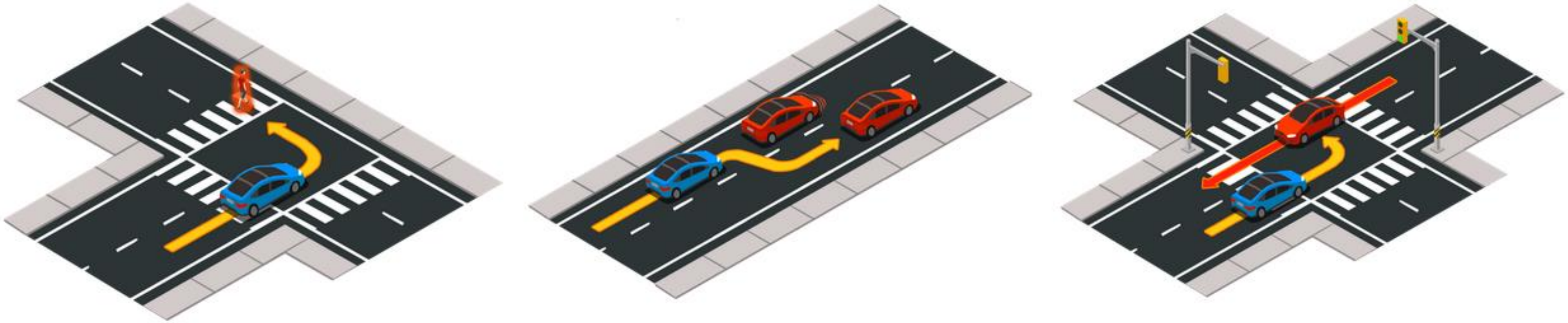


5. Autonomous Driving Stack



5. ADS Validation in CARLA Leaderboard

- Evaluate the **driving proficiency** of autonomous agents in **realistic traffic situations**
- **Open platform** for the community to perform **fair and reproducible evaluations**
- Autonomous agents have to drive through a set of **predefined routes** facing multiple traffic situations based on the **NHTSA typology**



- Two participation modalities: **SENSORS** and **MAP**

5. ADS Validation in CARLA Leaderboard

- **Metrics**
  - **Route Completion (RC)**: Percentage of route completed by the agent in a every route *i*, averaged across N routes

$$RC = \frac{1}{N}\sum_{i}^{N} R_i$$

  - **Infraction Penalty (IP):** Geometric series of infraction penalty coefficients, p_i^j for every instance *i* of infraction *j* incurred by the agent during the route.

$$P_i = \prod_{j}^{ped,\ldots,stop} (p_i^j)^{infractions(j)}$$

  - **Driving Score (DS)**: Weighted average of the infraction penalty with route completion for each route.

$$DS = \frac{1}{N}\sum_{i}^{N} R_i P_i$$

5. ADS Validation in CARLA Leaderboard

- **Results**

Local experiments setup

- **42 routes** from **6 different CARLA towns** (01 to 06)
- **3 repetitions** for each routes
- **7 weather conditions**
- **6 daylight conditions**

TABLE III: Local experiment B: We compare our ADS against different End-to-End architectures, showing the μ and σ over 3 evaluations for each model. We bold the best results in **black** and the second best in **blue** for each metric

Method	**Aux. Sup.**	**DS** ↑ [%]	**RC** ↑ [%]	**IP ↑ [0,1]**
CIRLS [27]	Velocity	22.97±0.90	35.46±0.41	0.66±0.02
LBC [28]	BEV Sem	29.07±0.67	61.35±2.26	0.57±0.02
AIM [29]	None	51.25±0.17	70.04±2.31	0.73±0.03
	2D Sem	57.95±2.76	80.21±3.55	0.74±0.02
AIM-MT	BEV Sem	60.62±2.33	77.93±3.06	0.78±0.01
	Dth+2D Sem	**64.86±2.52**	**80.81±2.47**	**0.80±0.01**
AIM-VA	2D Sem	60.94±0.79	75.40±1.53	0.79±0.02
NEAT [26]	BEV Sem	**65.10±1.75**	79.17±3.25	**0.82±0.01**
Ours	Modular	62.91±1.96	**92.11±1.84**	0.69±0.01

CARLA Leaderboard setup

- **100 routes evaluated in 2 secret CARLA towns**
  10 routes x 2 weather conditions x 5 repetitions
- **173 km of driving experiences**

TABLE IV: CARLA Autonomous Driving Leaderboard results (MAP track). We bold the best results in **black** and the second best in **blue** for each metric

Team	Submission	**DS** ↑ [%]	**RC** ↑ [%]	**IP** ↑ [0,1]
Anonymous	GRI-based DRL [30]	**33.78**	57.44	**0.57**
RobeSafe	Techs4AgeCar (Ours)	**18.75**	**75.11**	0.28
ERDOS	Pylot [10]	16.70	48.63	**0.50**
LRM-B	CaRINA [2]	15.55	40.63	0.47
RobeSafe	SmartElderlyCar [4]	12.63	**61.59**	0.33

[Gómez-Huélamo et al, 2022] "How to build and validate a safe and reliable Autonomous Driving stack? A ROS based software modular architecture baseline", C. Gómez-Huélamo, A. Diaz-Diaz, J. Araluce, M.E. Ortiz, R. Gutiérrez, F. Arango, A. Llamazares and L.M. Bergasa, **IEEE Intelligent Vehicles Symposium (IV) 2022,** Aachen, Germany, June 2022. Accepted for publication

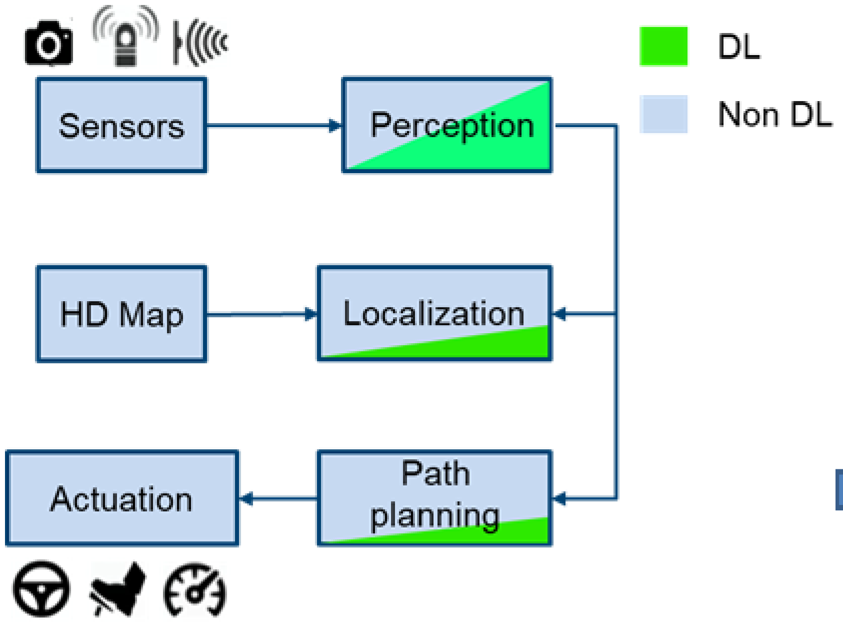
5. ADS Validation in CARLA Leaderboard

[Díaz-Díaz et al, 2022] "HD maps: Exploiting OpenDRIVE potential for Path Planning and Map Monitoring", A. Díaz-Díaz, M. Ocaña, A. Llamazares, C. Gómez-Huélamo, Pedro Revenga and L.M. Bergasa, **IEEE Intelligent Vehicles Symposium (IV) 2022,** Aachen, Germany, June 2022. Accepted for publication

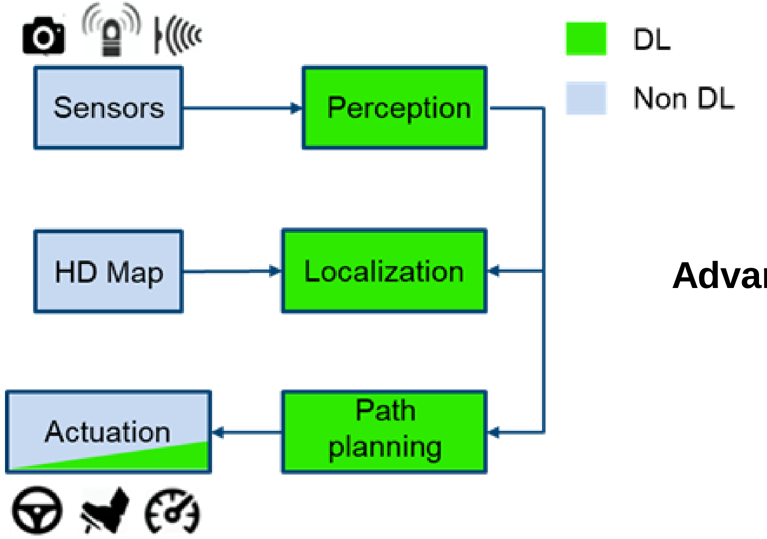
Conclusions

- **CARLA** simulator proves to be a **valid source of artificial data** for automotive industry
  - Synthetic data has a **good potential** for the training and validation of **DL perception models**
  - **More sophisticated sensor models are needed**
  - **Useful** for **end-to-end and reinforcement learning** model implementation
- **CARLA** is a **good tool to evaluate driving proficiency** of AD agents
  - Open Leaderboard
  - Fair and reproducible evaluations

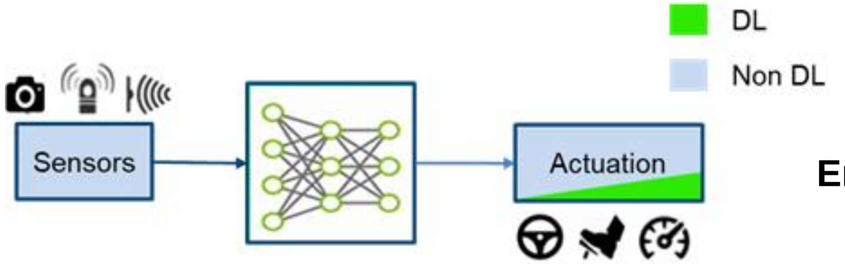
Next Steps



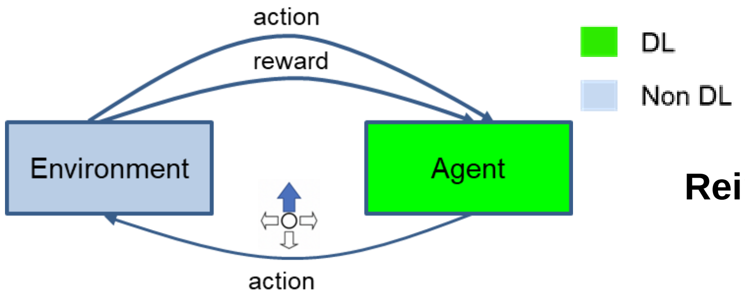
Classical AD Stack



Advanced AD Stack



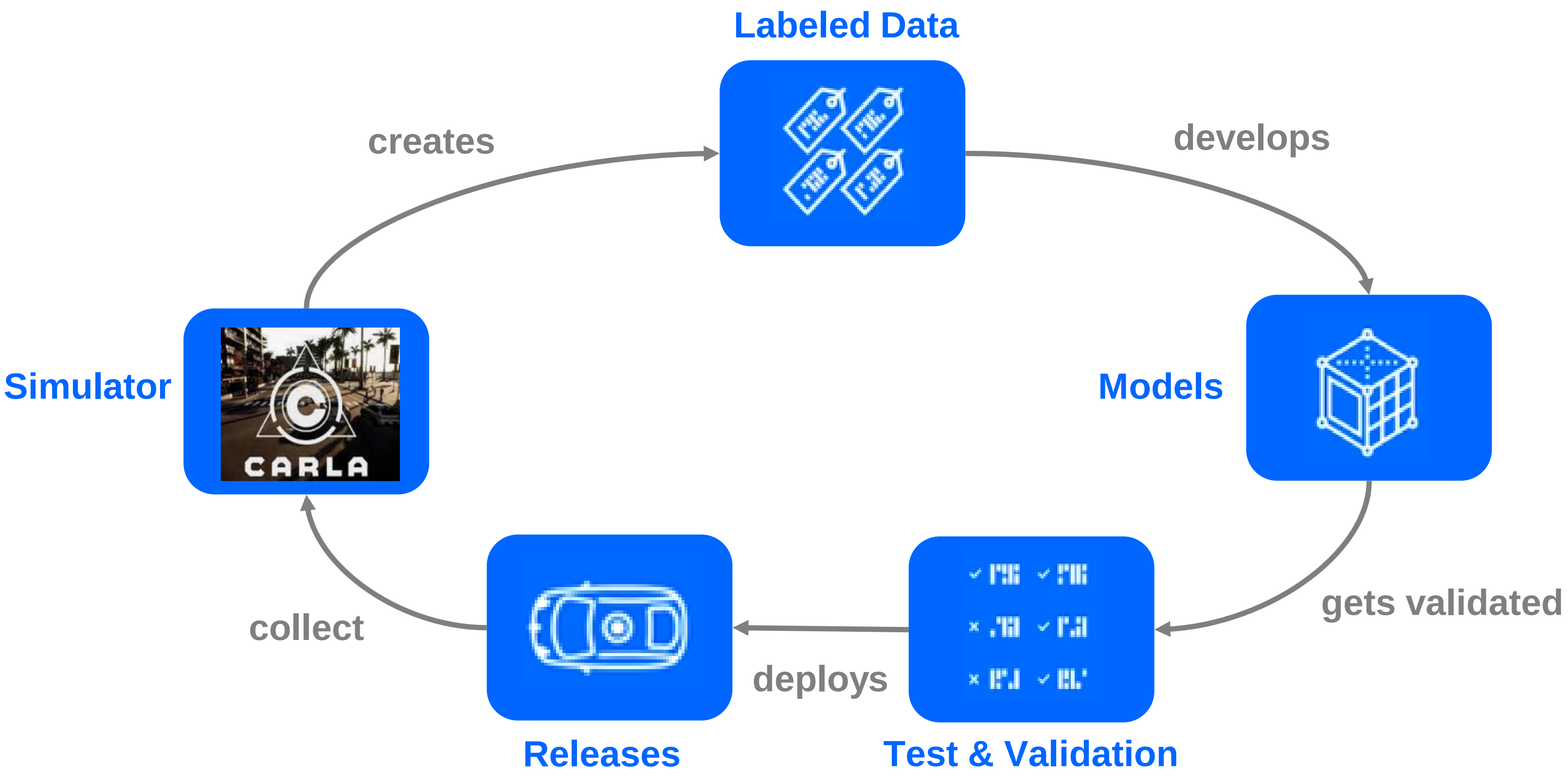
End-to-end AD



Reinforcement Learning AD

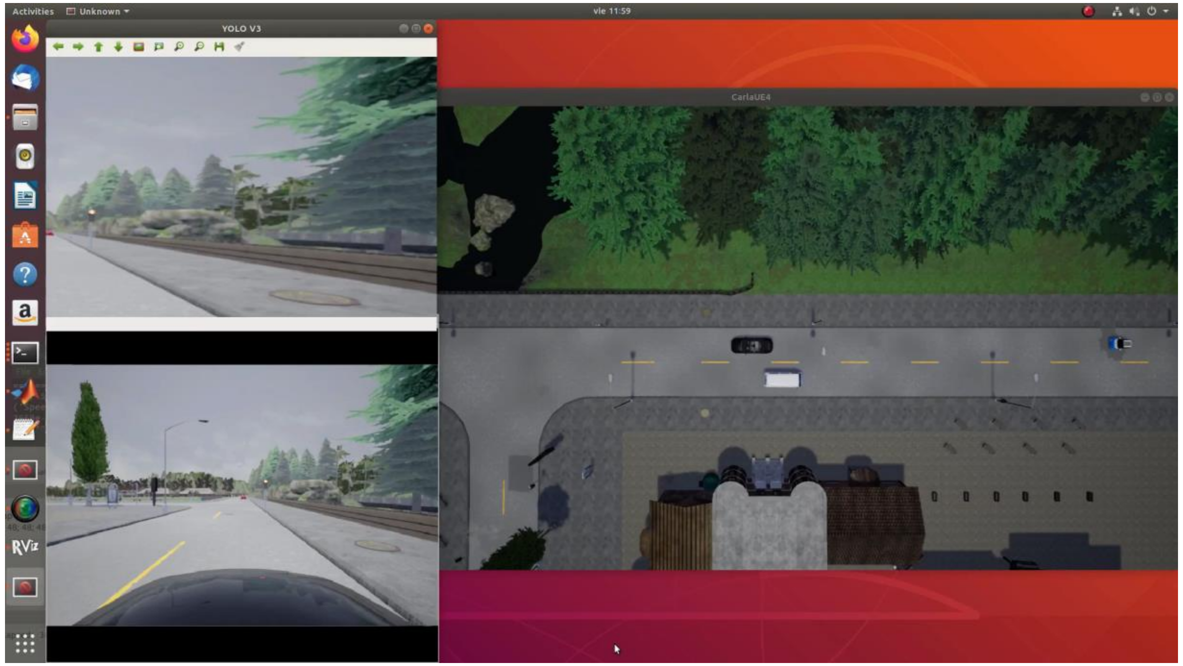
Next Steps

- Make CARLA become a factory for training and validating AD



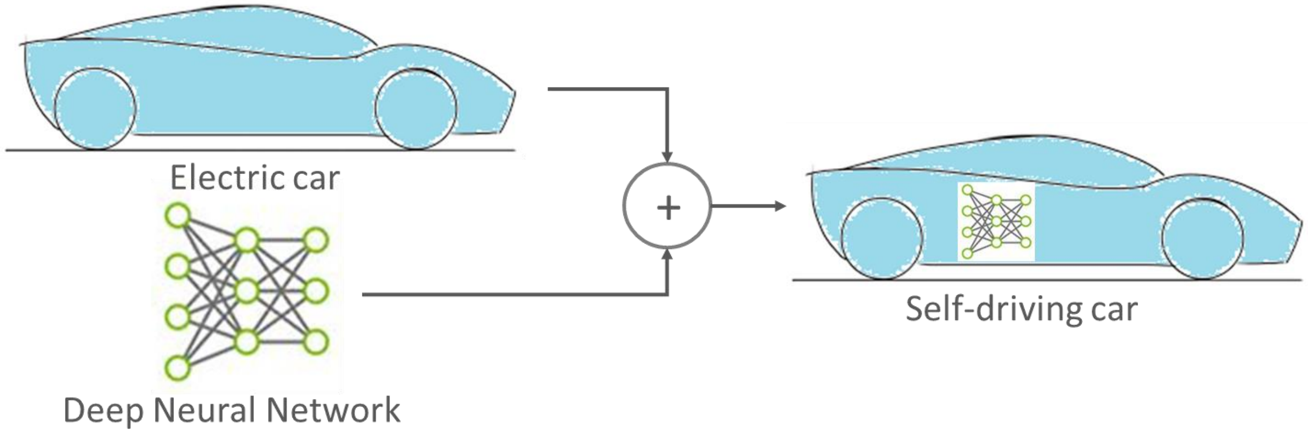
Next Steps

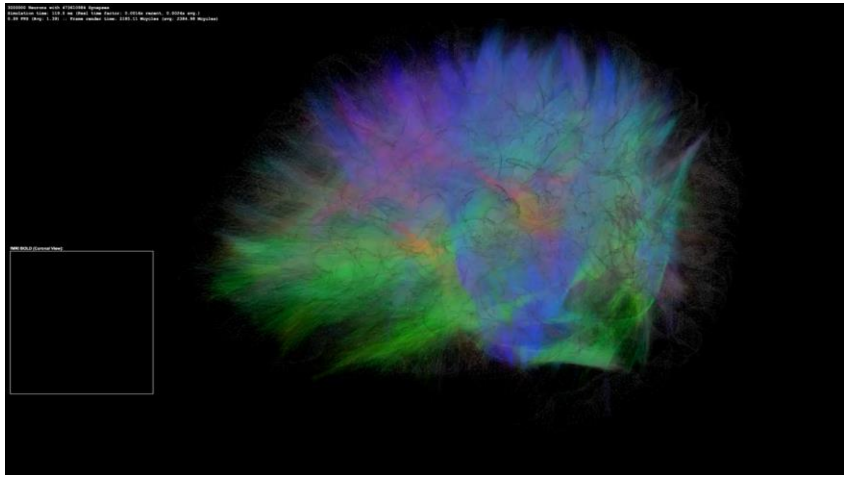
- Improve simulation and large-scale testing



- More realistic sensors
- More realistic drivers and pedestrian behaviors
- Learn agent models from real world
- Use logs from driving and create variations (GAN, Transformers, etc.)
- Scale to different cities and different countries

Final conclusion





“The future of the DL depends on some graduate student who is deeply suspicious of everything I have said.”

Geoffrey Hinton

“Godfather of Deep Learning”

THANKS!

luism.bergasa@uah.es

www.robesafe.uah.es/personal/bergasa/

Universidad de Alcalá

RobeSafe