P.A.I.S.S. PRAIRIE ARTIFICIAL INTELLIGENCE SUMMER SCHOOL

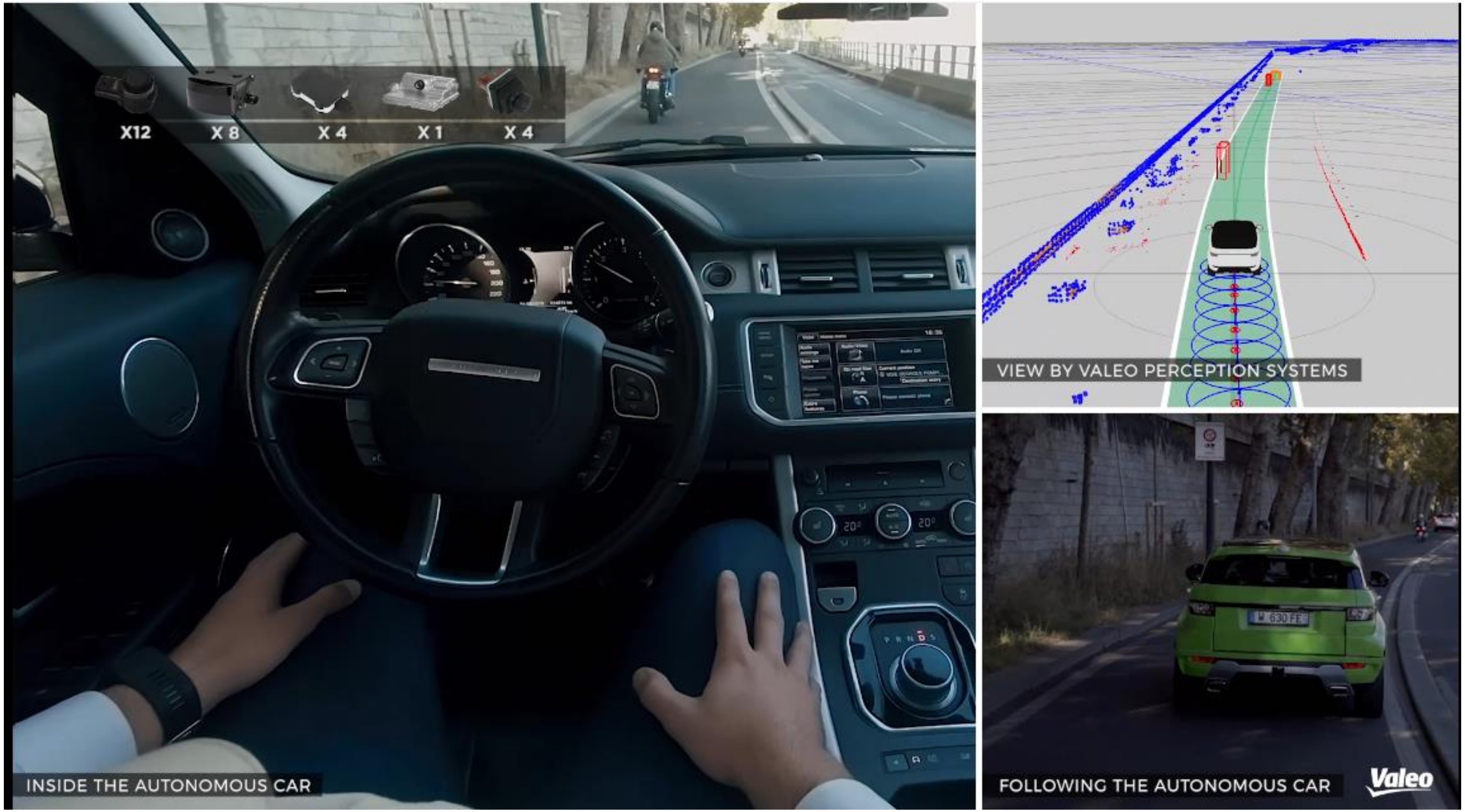
3 >5 OCTOBER 2019 - PARIS

Autonomous Driving
& Sustainable Supervision

Patrick Pérez

Valeo
SMART TECHNOLOGY FOR SMARTER CARS

valeo.ai



https://www.youtube.com/watch?v=vE0h3Yy458k

Promises & Challenges

Saving lives, time, energy

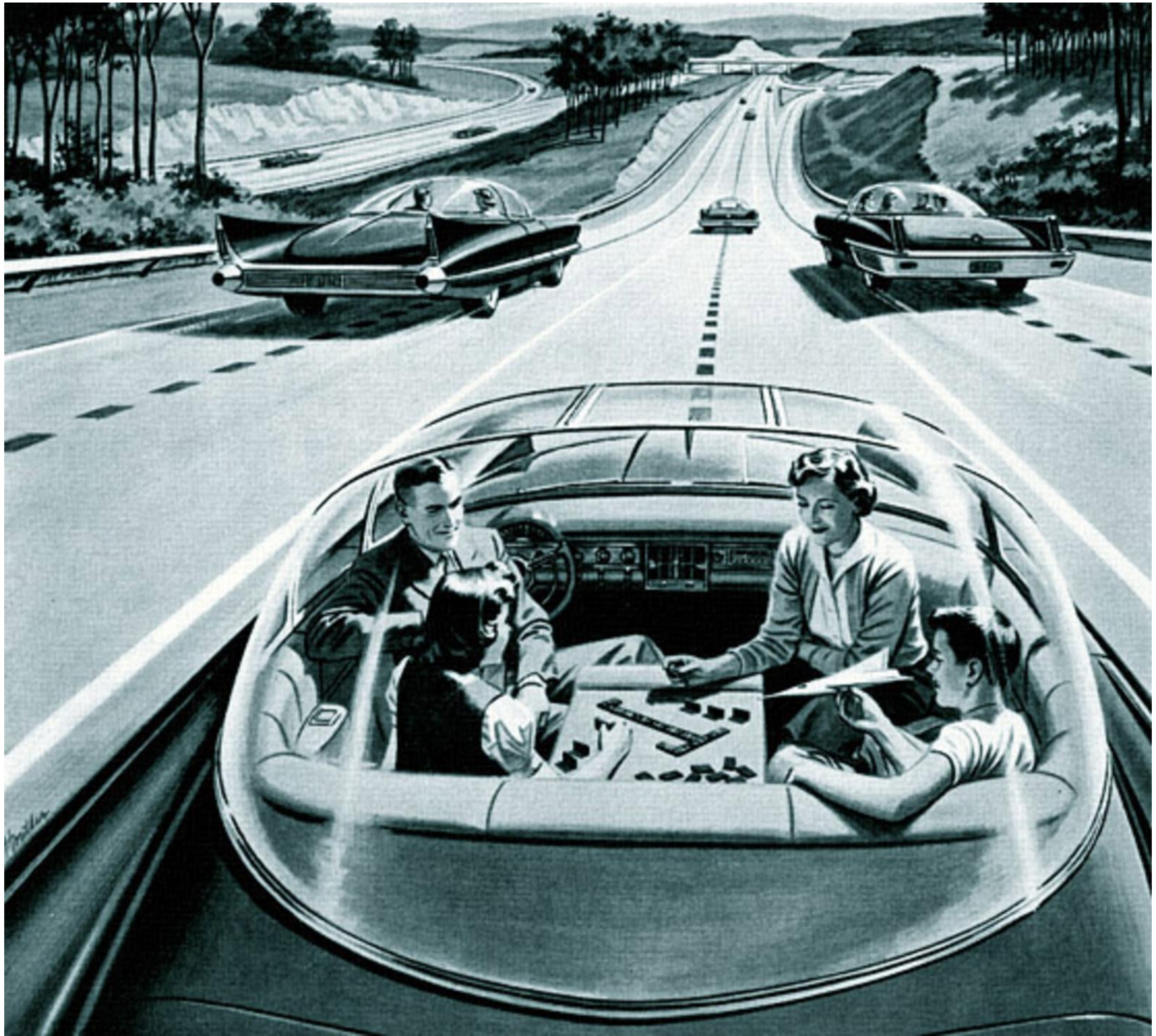
Safety critical, real-time, embedded AI in the wild

Driving AI must be:

accurate, robust, able to generalize well

validated, certified, explainable

Driverless Cars



source

Driverless Cars

Eureka-Prometheus 86-95

https://www.youtube.com/watch?v=I39sxwYKlEE

CVPR 2019

Driverless Cars

Automation Levels

0 No Assistance

Level 0

"No automation"

The driver controls everything.

1 Driver Assistance

Level 1

"Feet off"

The car controls certain functions like accelerating and braking.

2 Partial Automation

Level 2

"Hands off"

Both steering and acceleration/ deceleration is automatically controlled but the driver must supervise at all times.

3 Conditional Automation

Level 3

"Eyes off"

The driver can engage in secondary tasks while the car drives itself, but the driver must be able to take back control whenever requested.

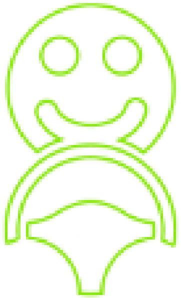
4 High Automation

Level 4

"Mind off"

The driver becomes a passenger as the car full takes over for a part or the entire journey (may include driverless operation).

5 Full Automation

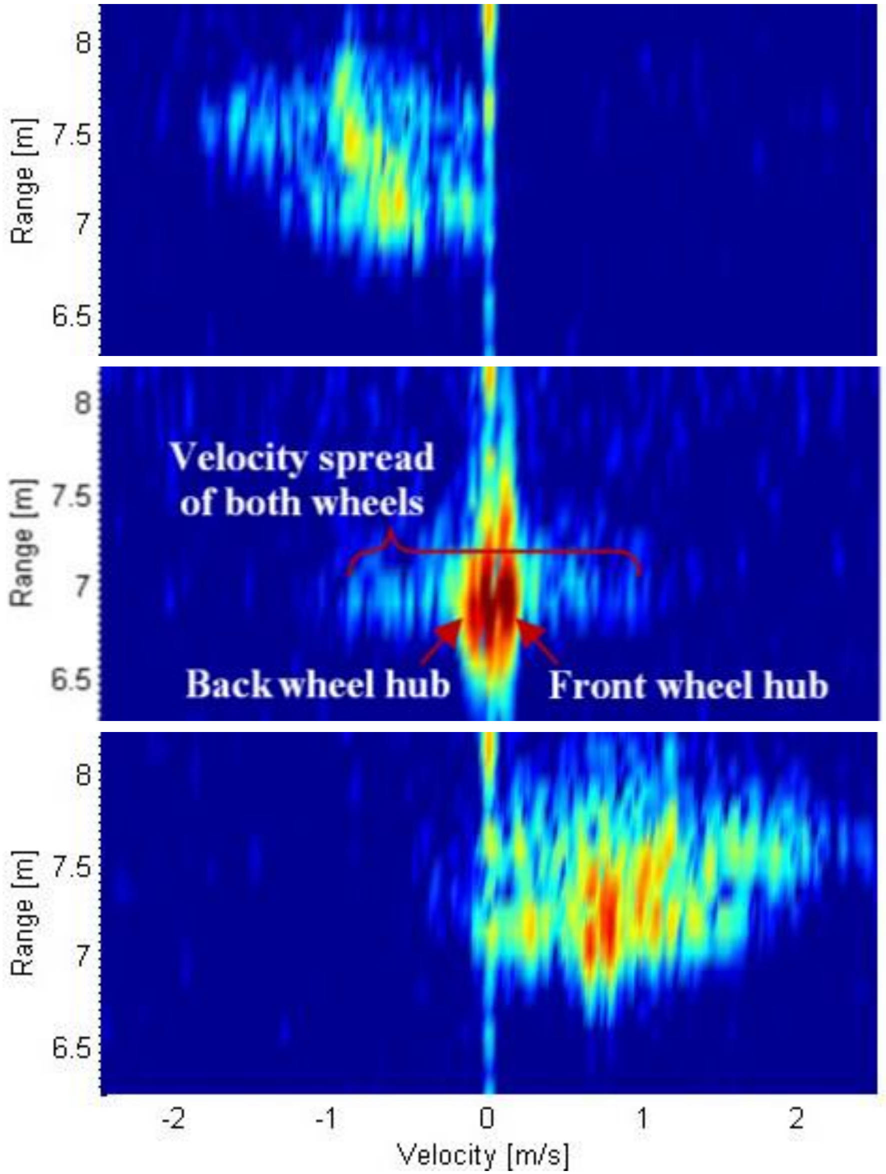


Level 5

"Driverless"

The car autonomously handles all situations normally controlled by a human driver.

Radar Range-Doppler plots



[Shubert 2015]

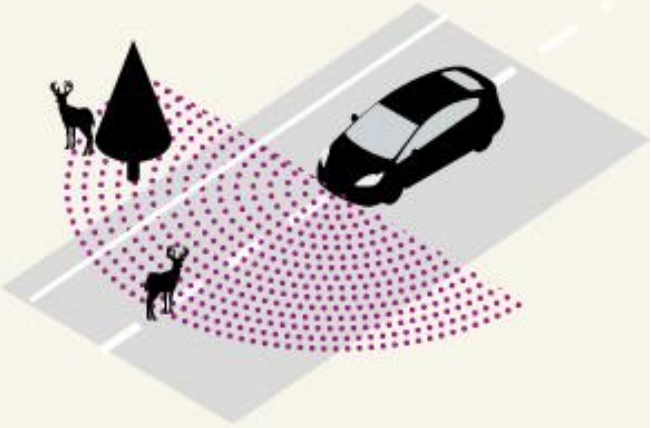
Key sensors

● **Camera**

Takes images of the road that are interpreted by a computer. Limited by what the camera can "see".

● **Radar**

Radio waves are sent out and bounced off objects. Can work in all weather but cannot differentiate objects

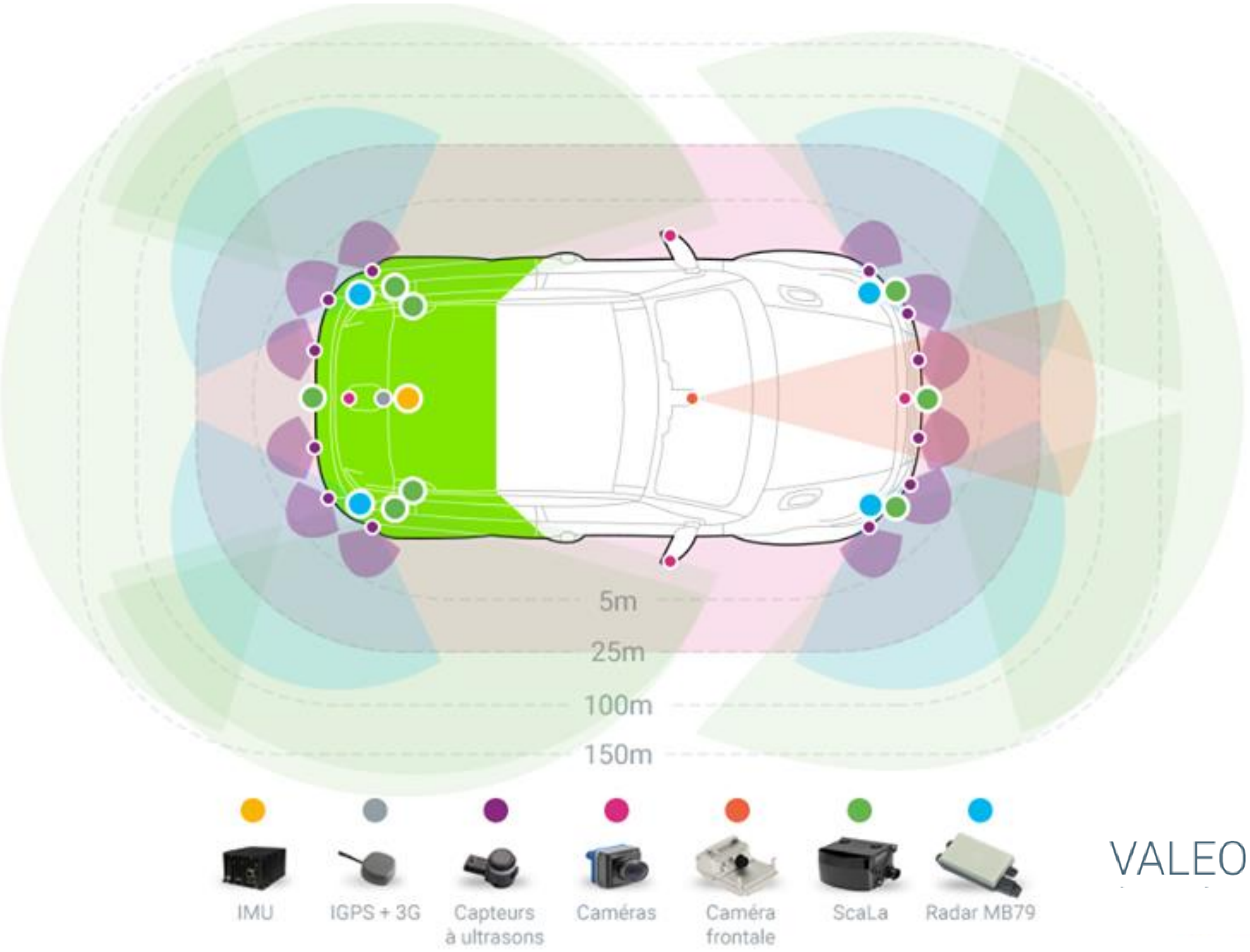


● **LiDAR**

Light pulses are sent out and reflected off objects. Can define lines on the road and works in the dark.

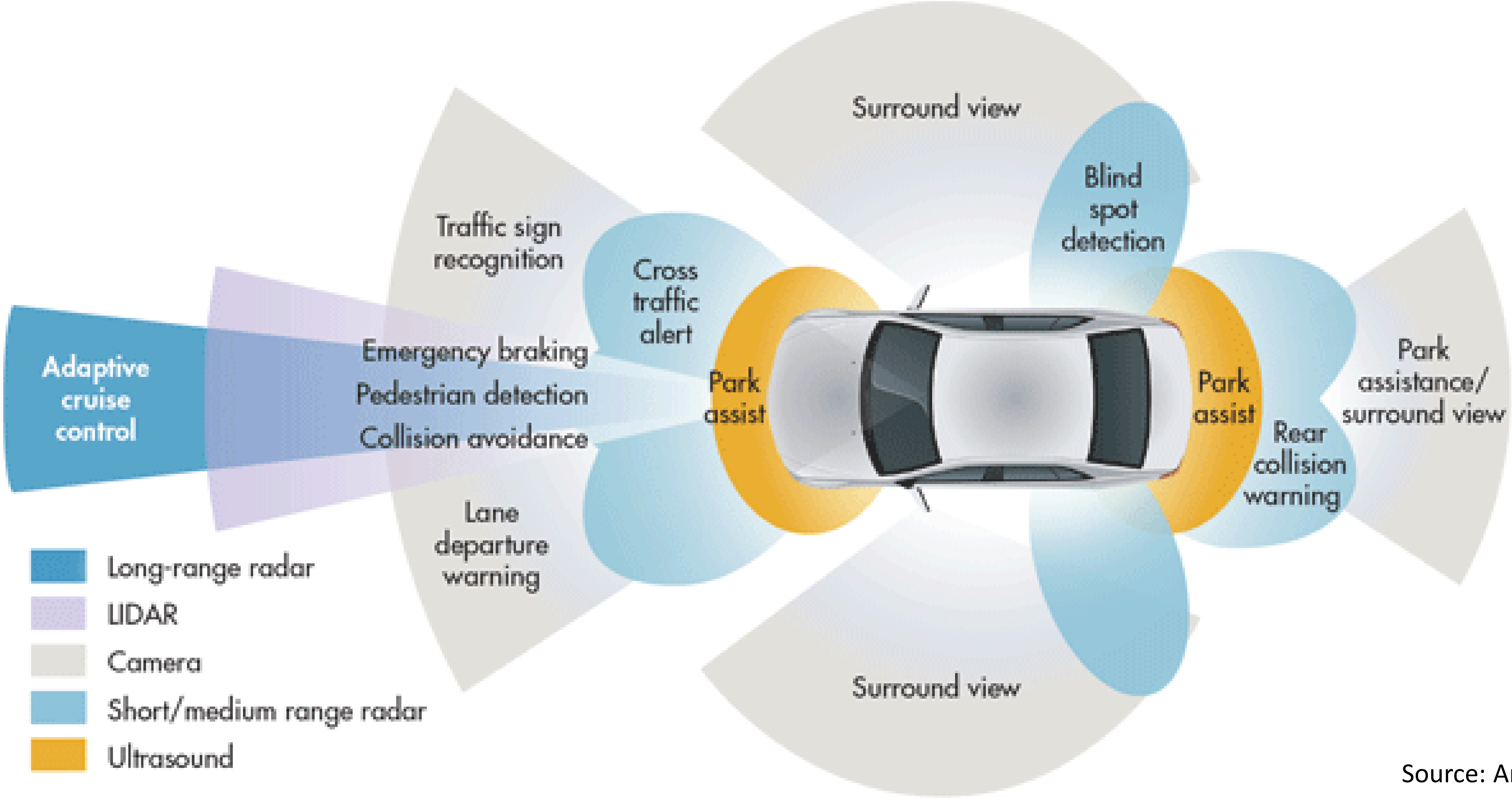
Source: Delphi
Reuters/©Gulf News

Sensor suite



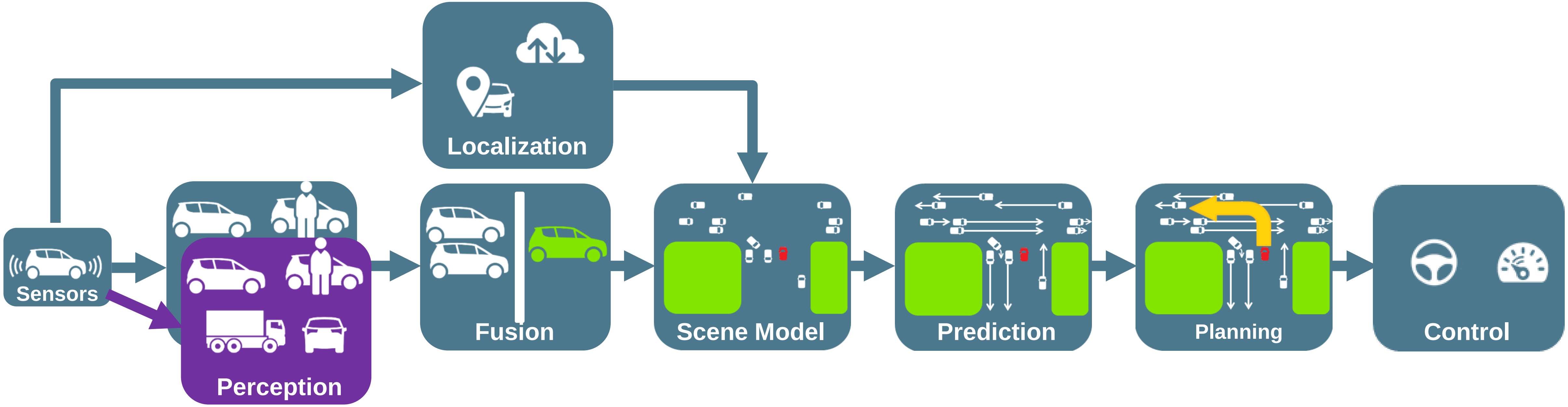
VALEO DRIVE4U®

Advanced Driving Assistance Systems (ADAS)



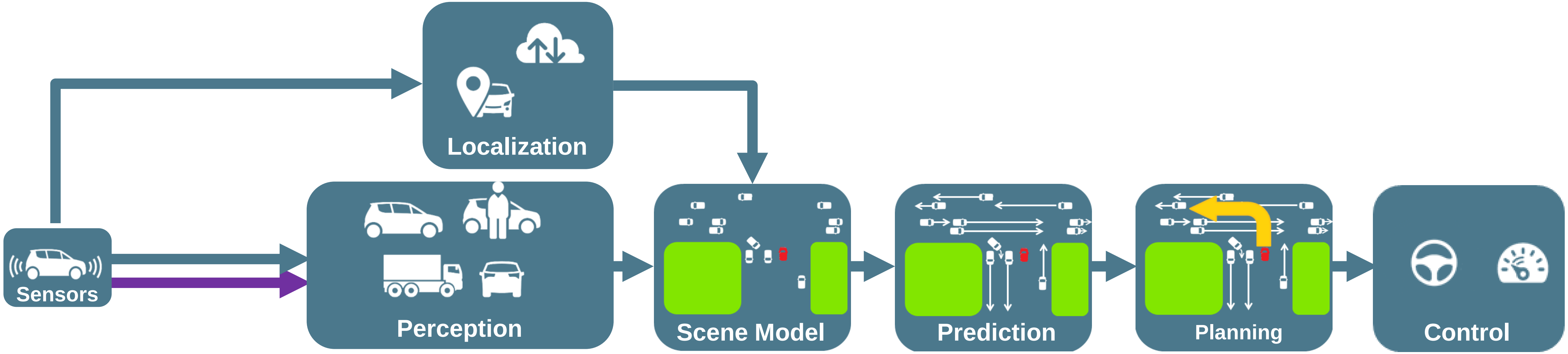
Source: Ansys

Autonomous Driving (AD) Systems



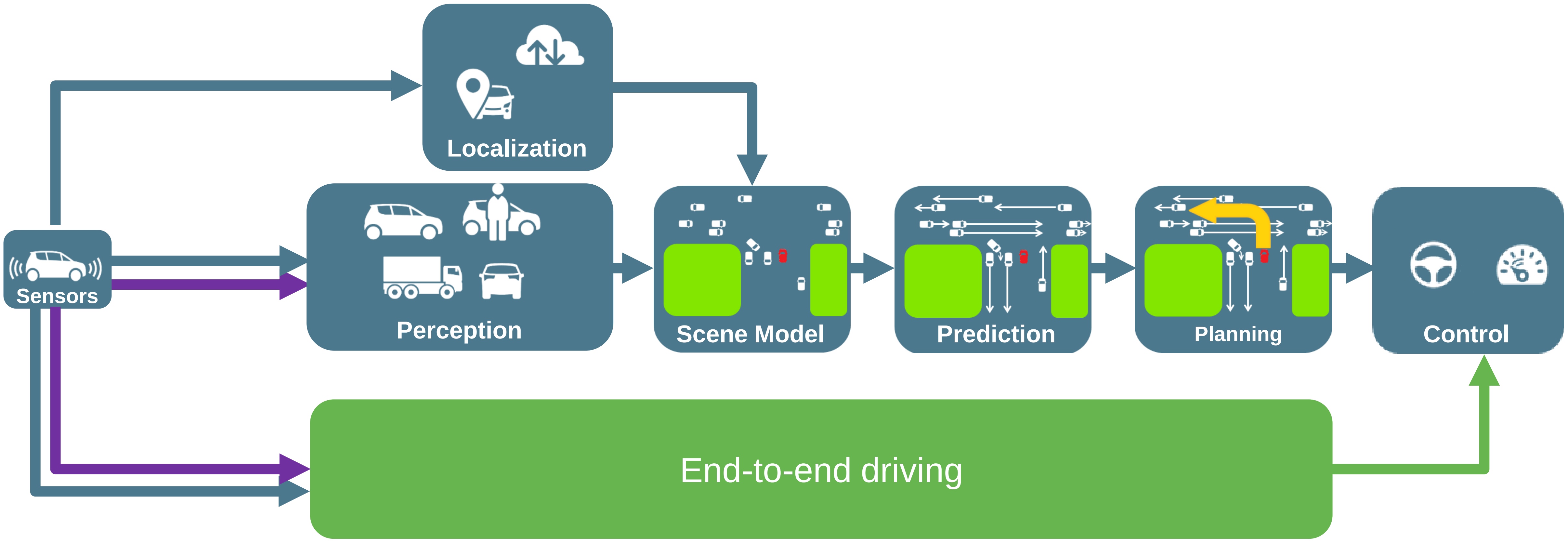
Courtesy X. Perrotton@valeo

Autonomous Driving (AD) Systems



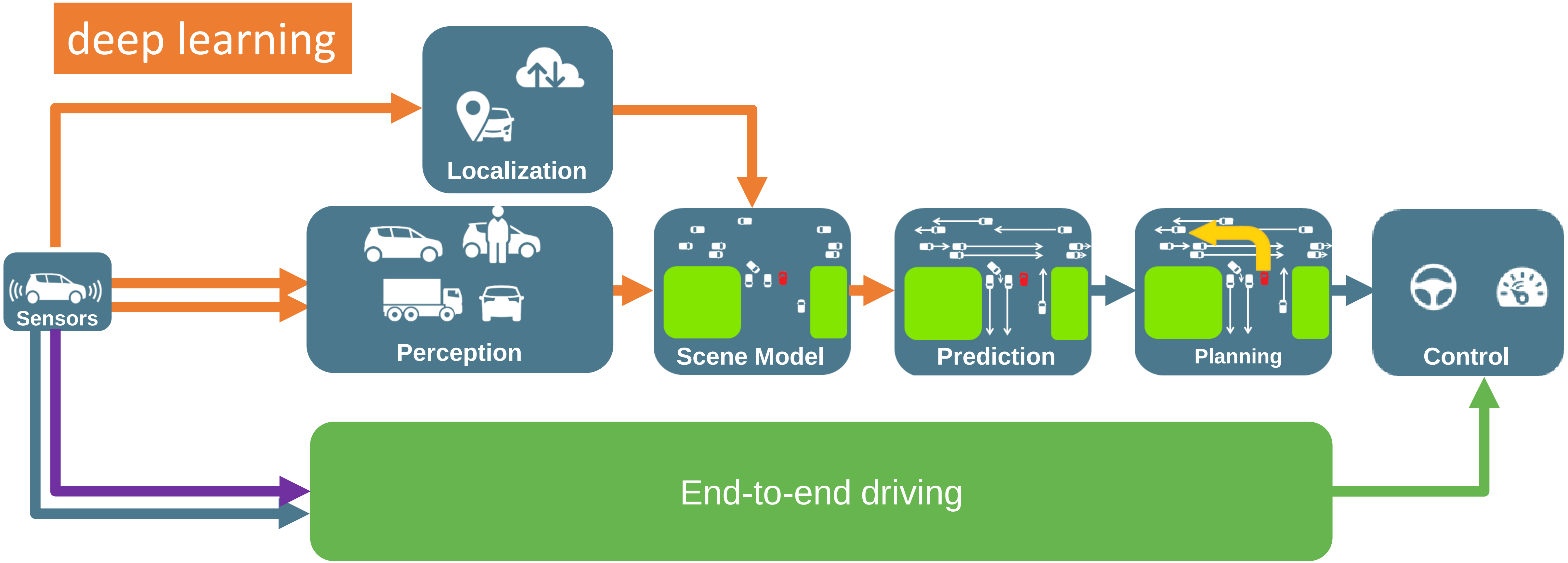
Courtesy X. Perrotton@valeo

Autonomous Driving (AD) Systems



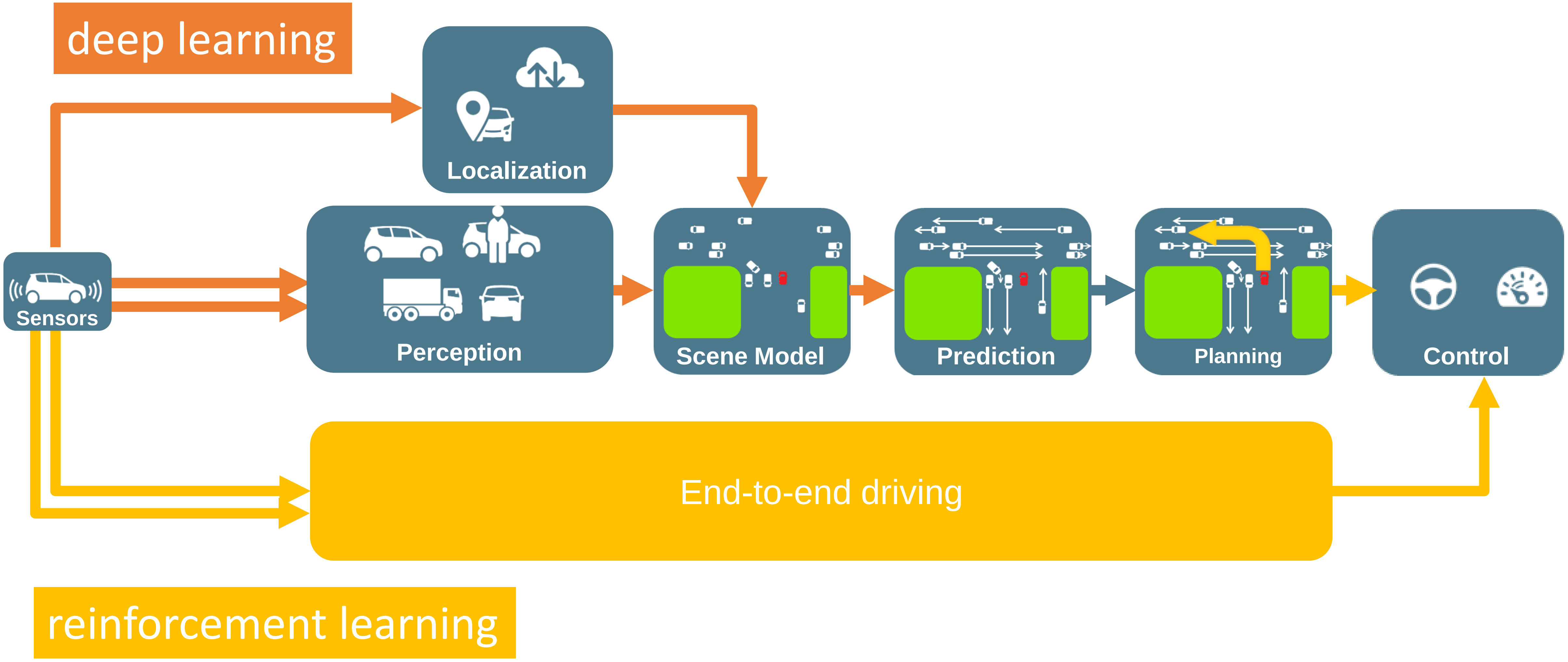
Courtesy X. Perrotton@valeo

Autonomous Driving (AD) Systems



Courtesy X. Perrotton@valeo

Autonomous Driving (AD) Systems



Courtesy X. Perrotton@valeo

2D/3D Scene Understanding

Detect (bounding boxes with categories)

- Vehicles , vulnerable road users (VRUs), signs, road work

Segment (pixel/point labelling)

- Road, pavement, free/drivable space, lane marks

Measure (pixel/point regression)

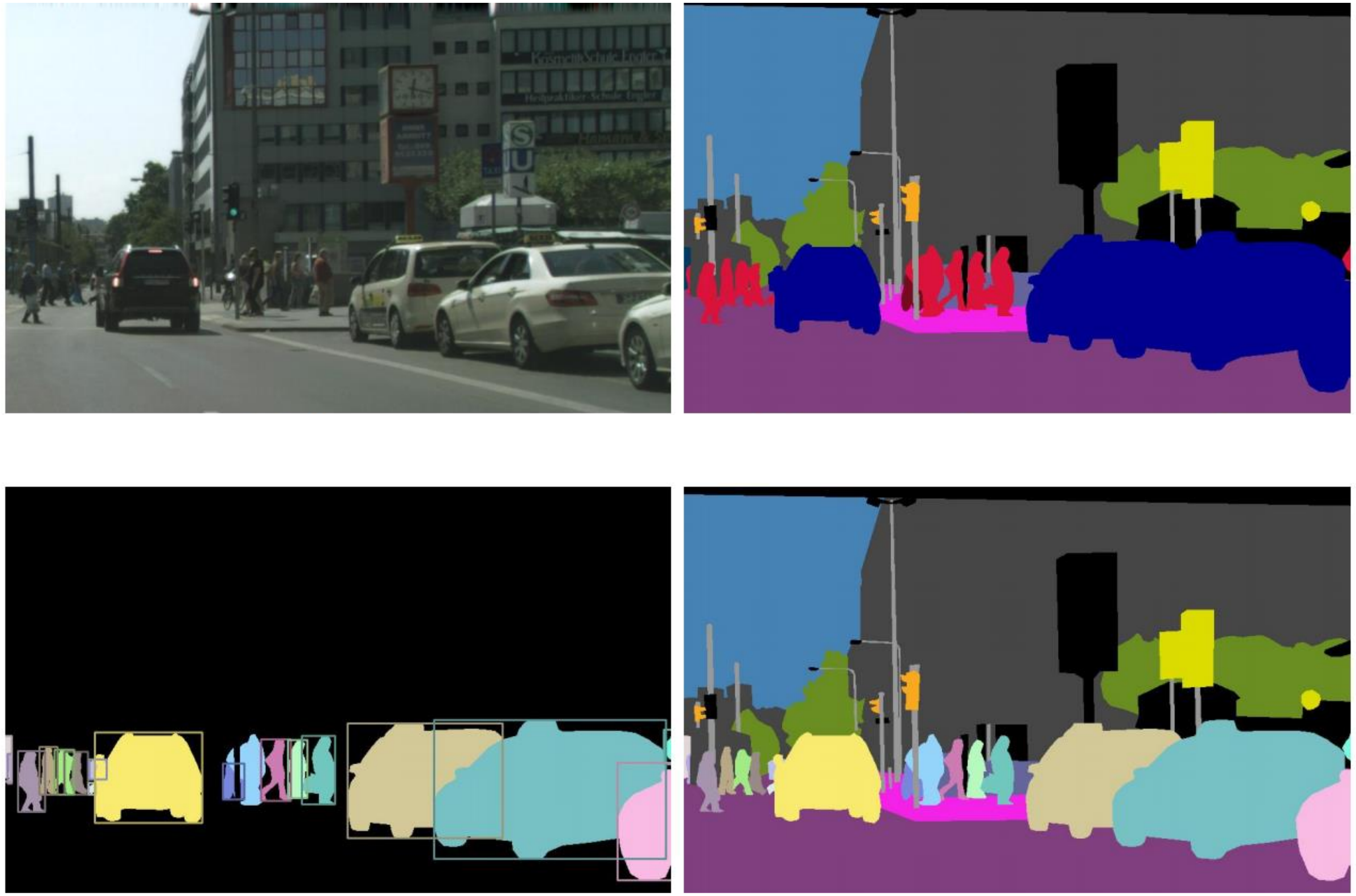
- Distance, speed

Analyze (object-level)

- Sub-categories, attributes, ‘intention, ‘attention’, next position

2D Semantic Segmentation

Variants: Semantic, instance, plenoptic



2D Semantic Segmentation

Variants: Semantic, instance, plenoptic

Metric: Mean intersection over union (mIoU)

2D Semantic Segmentation

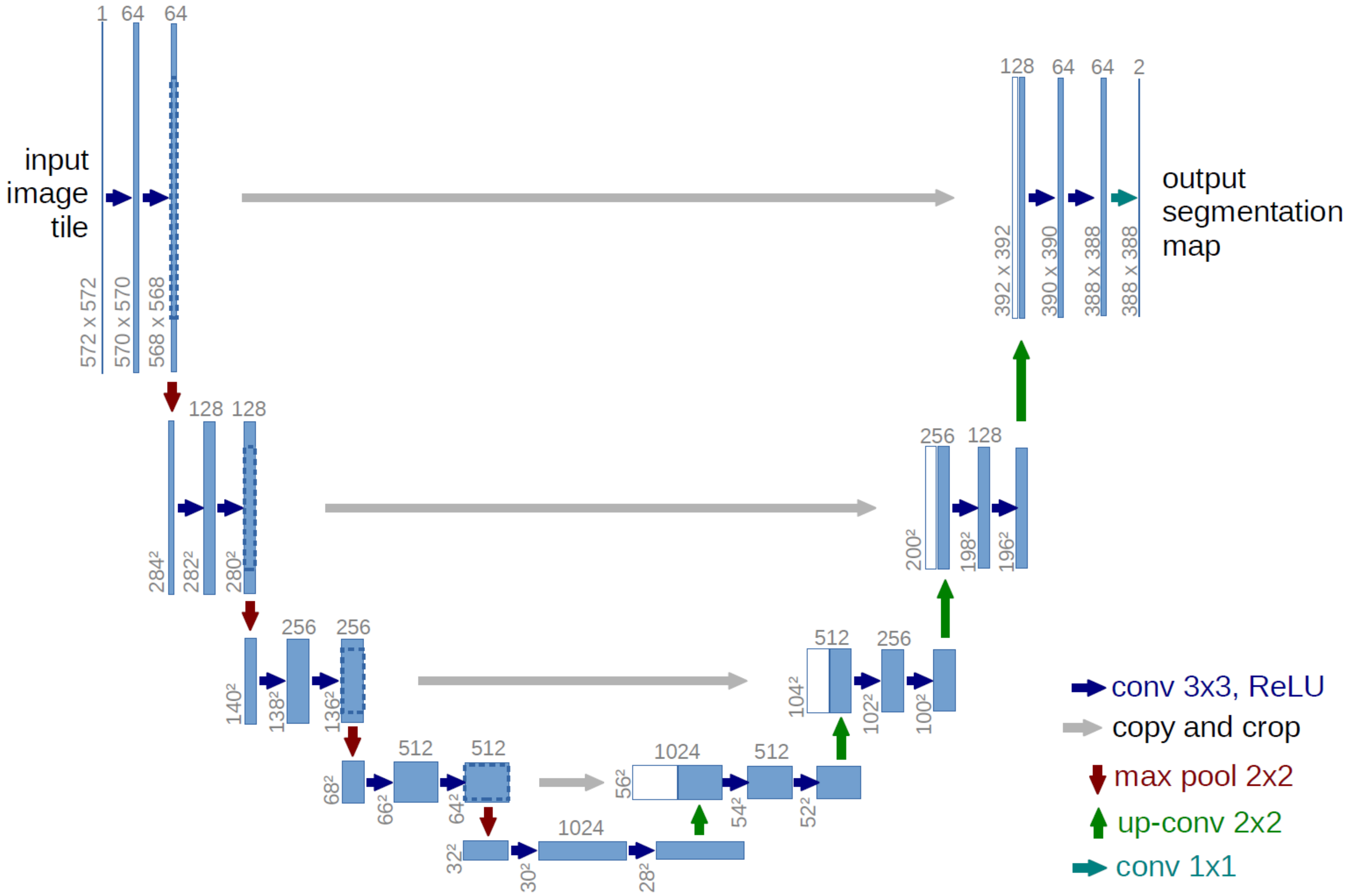
Variants: Semantic, instance, plenoptic

Metric: Mean intersection over union (mIoU)

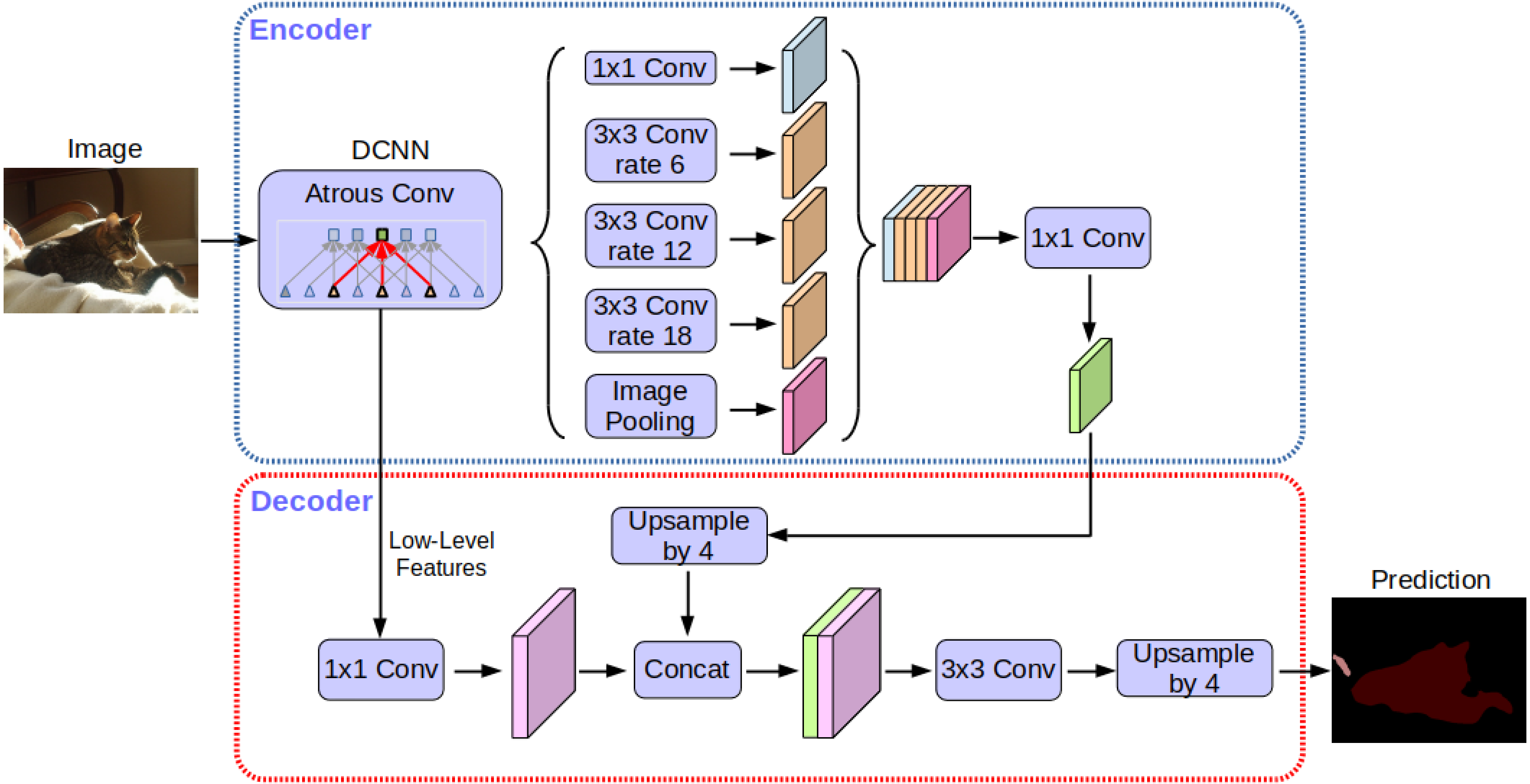
Key architecture: Fully convolutional encoder-decoder

- Favorite deep ConvNet as backbone encoder
- Skip connections à la U-net for full res decoding

U-net (2015)


1
64
64
input
image
tile
572 x 572
570 x 570
568 x 568
128
128
284²
282²
280²
256
256
140²
138²
136²
512
512
68²
66²
64²
1024
32²
30²
28²
1024
512
56²
54²
52²
512
256
104²
102²
100²
256
128
200²
198²
196²
128
64
64
2
392 x 392
390 x 390
388 x 388
388 x 388
output
segmentation
map
conv 3x3, ReLU
copy and crop
max pool 2x2
up-conv 2x2
conv 1x1

DeepLab-v3 (2018)

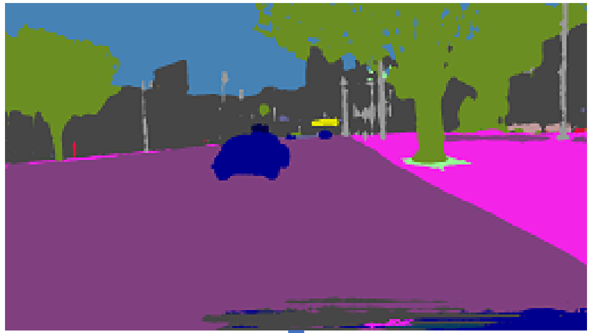


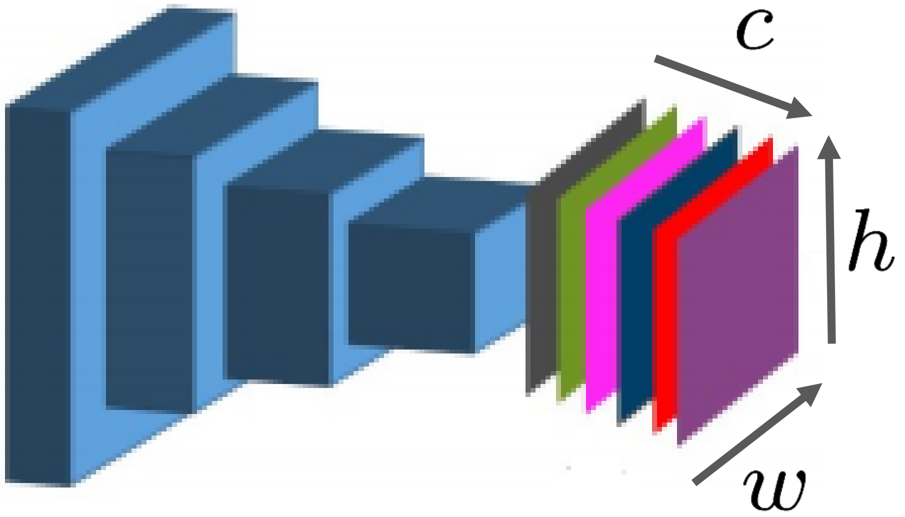
DeepLab-v3 (2018)

Training Semantic Segmentation

$\boldsymbol{x}$

$\boldsymbol{y}$





$\boldsymbol{P}_{\boldsymbol{x}}$ “soft” segmentation map after softmax

pixel-level cross entropy

$$\mathcal{L}_{\text{seg}}(\boldsymbol{x}, \boldsymbol{y}) = -\sum_{h}\sum_{w}\sum_{c} \boldsymbol{y}[h, w, c] \log \boldsymbol{P}_{\boldsymbol{x}}[h, w, c]$$

Stochastic Gradient Descent on $\sum_{(\boldsymbol{x},\boldsymbol{y})\in\text{Train}} \mathcal{L}_{\text{seg}}(\boldsymbol{x}, \boldsymbol{y})$

Driving Datasets

- CamVid – ECCV 2008
- **KITTI – IJRR 2013**
- Cityscapes – CVPR 2016
- Oxford Robotcar – IJRR 2016
- BDD100K – CVPR 2017
- *Mapillary* Vistas – ICCV 2017
- ApolloScape (*Baidu*) – CVPR 2018
- HDD (*Honda*) – CVPR 2018

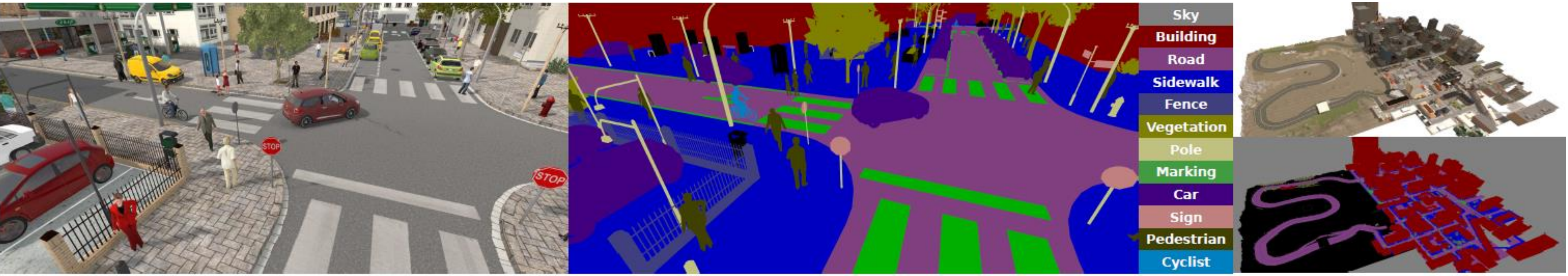
synthetic

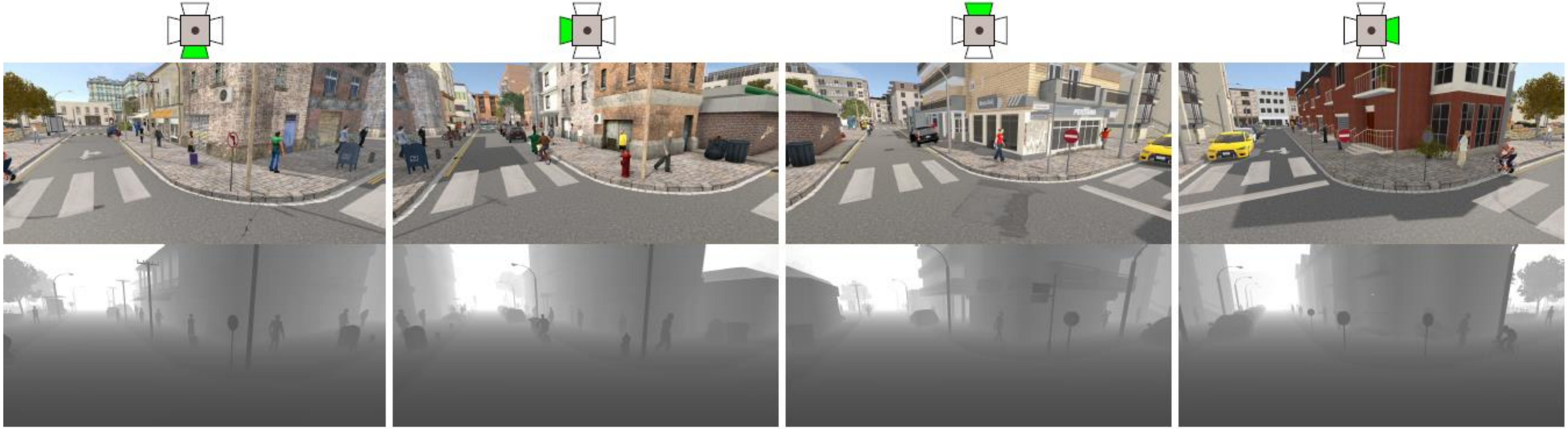
- GTA5 – ECCV 2016
- Synthia – CVPR 2016
- Carla simulator – arXiv 2017

2019

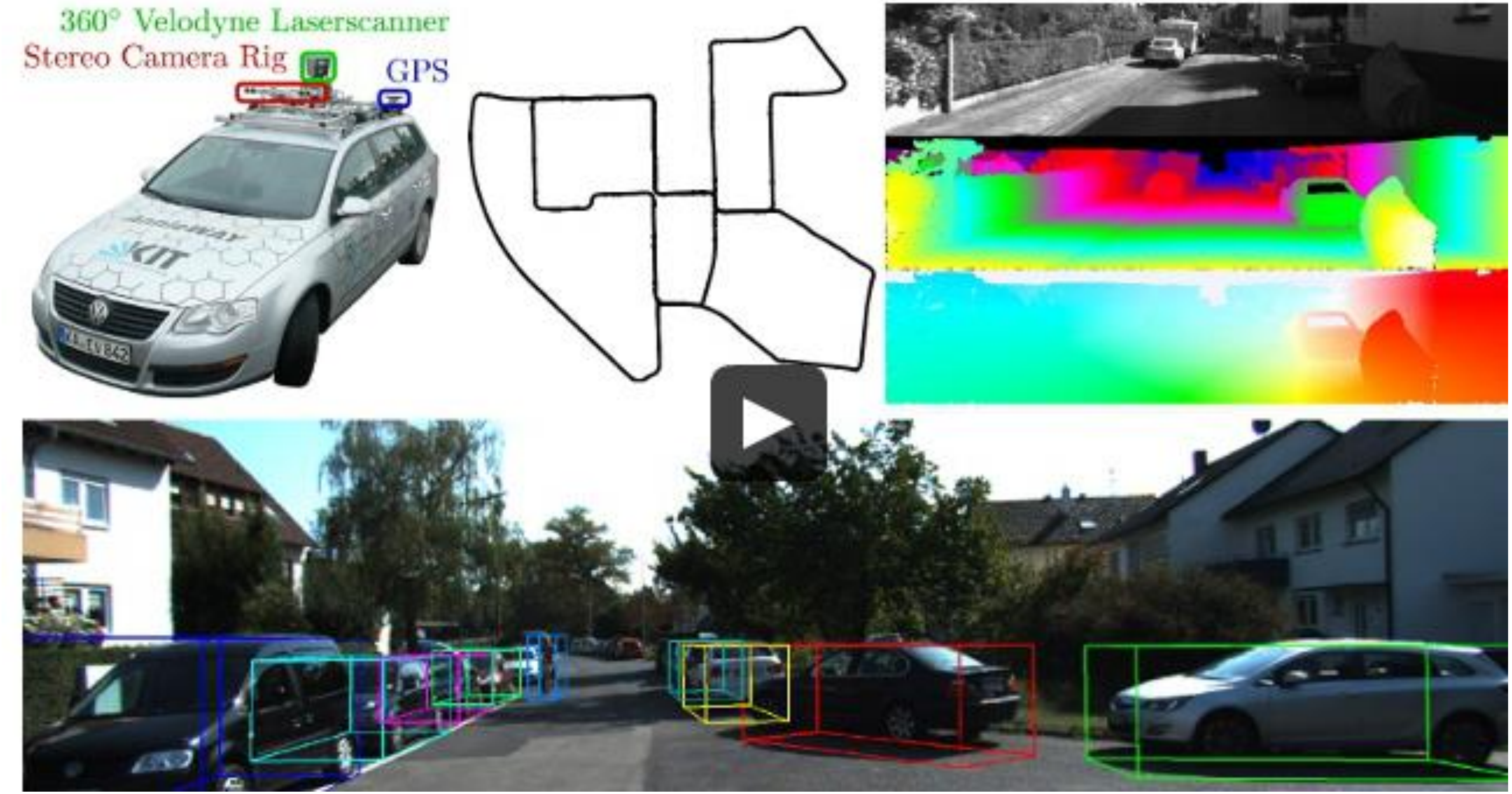
- India Driving Dataset – WACV 2019
- nuScenes (*Aptiv*) – arXiv 2019
- *Waymo* Open Dataset –2019
- *Lyft* Level 5 AV Dataset – 2019
- D2 City (*Didi*) – arXiv 2019
- A2D2 (*Audi*) – 2019
- Woodscape (*Valeo*) – ICCV 2019

Synthia


Sky
Building
Road
Sidewalk
Fence
Vegetation
Pole
Marking
Car
Sign
Pedestrian
Cyclist



KITTI



CITYSCAPES DATASET

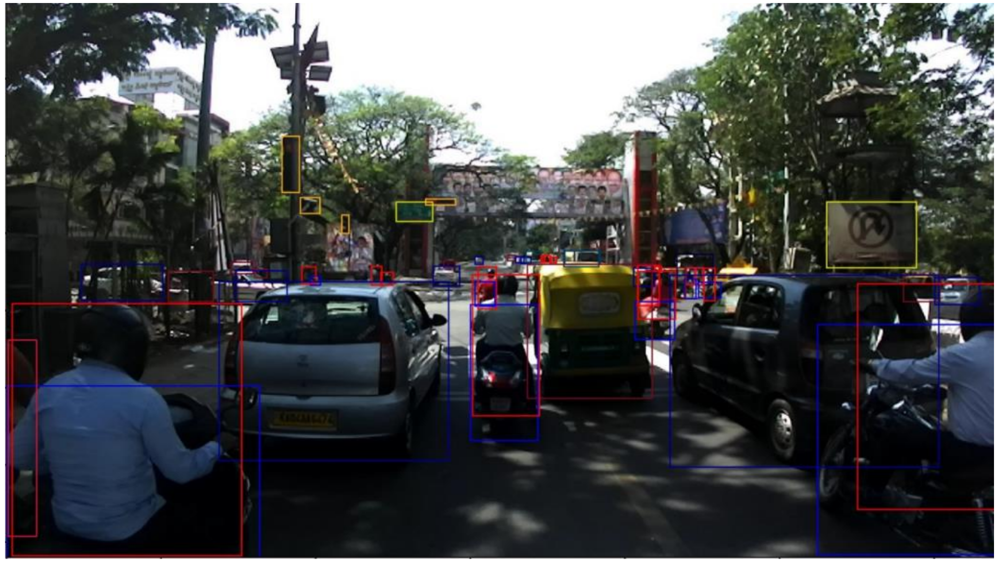
max planck institut informatik

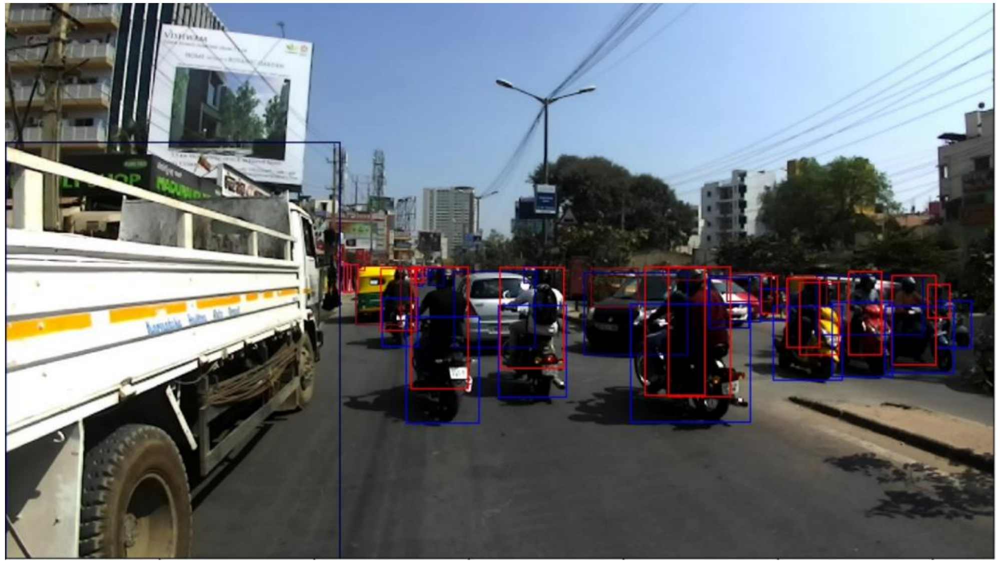
TECHNISCHE UNIVERSITÄT DARMSTADT

DAIMLER

IDD

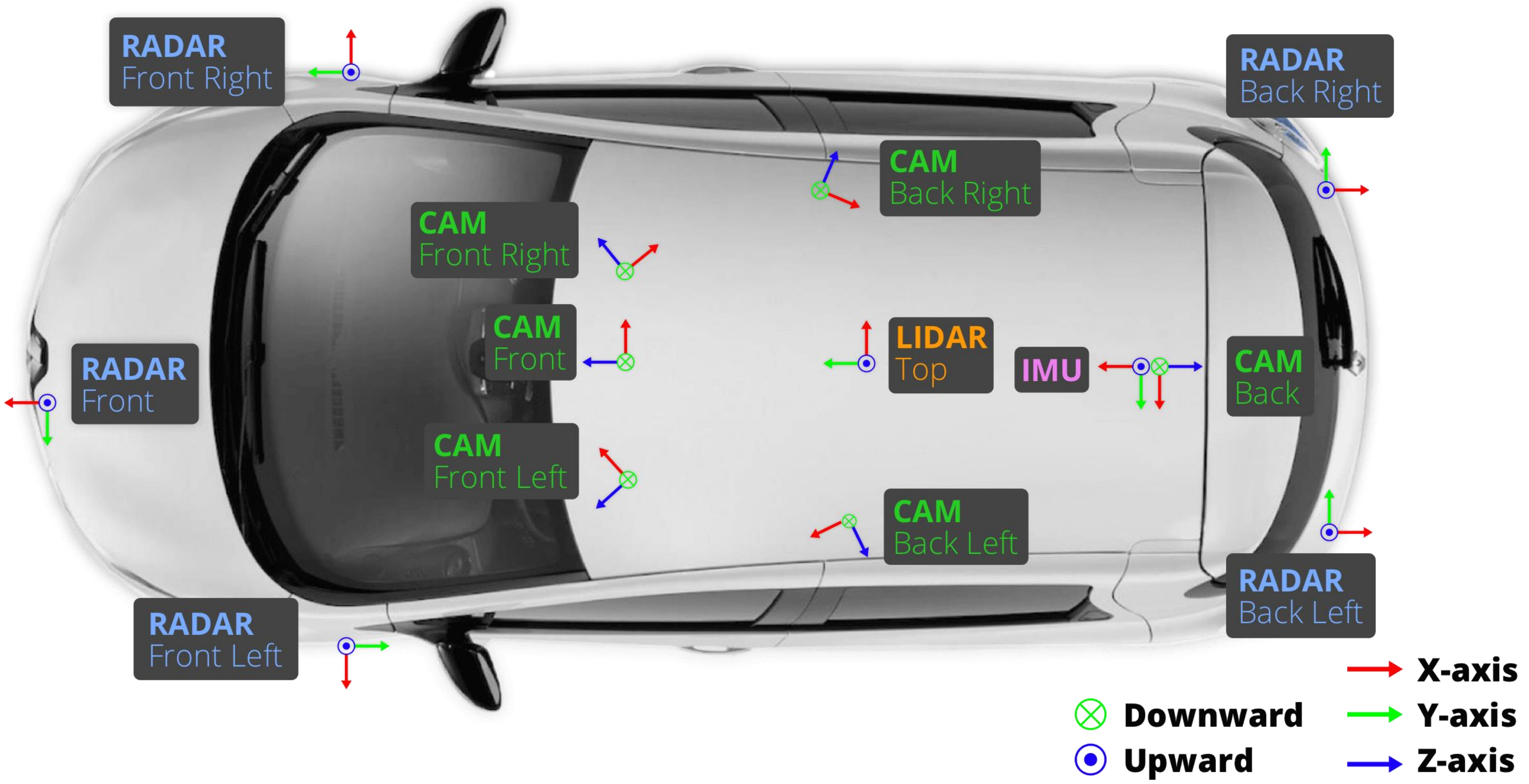
INTERNATIONAL INSTITUTE OF INFORMATION TECHNOLOGY
HYDERABAD





nuScenes

APTIV

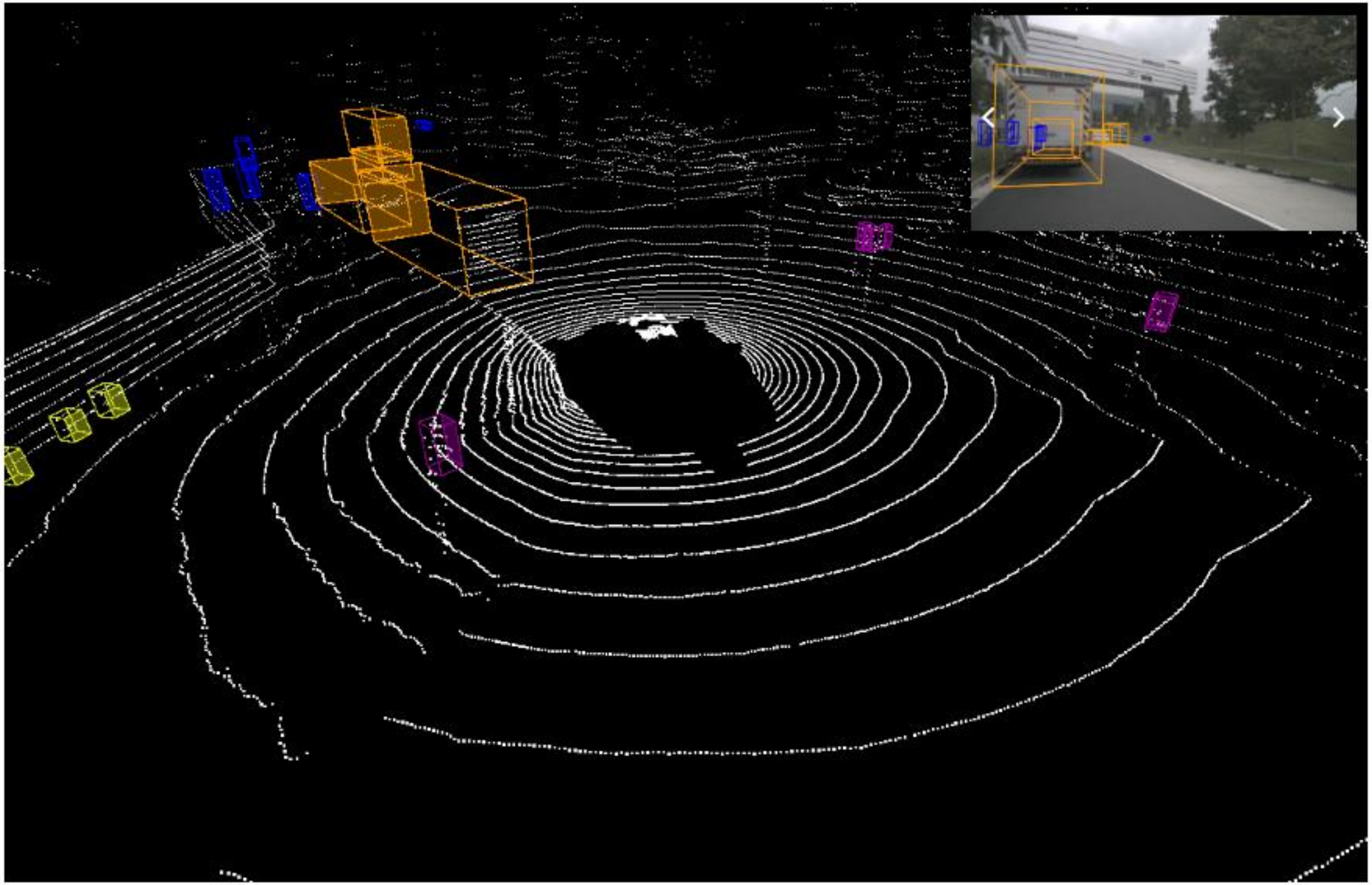


nuScenes

APTIV

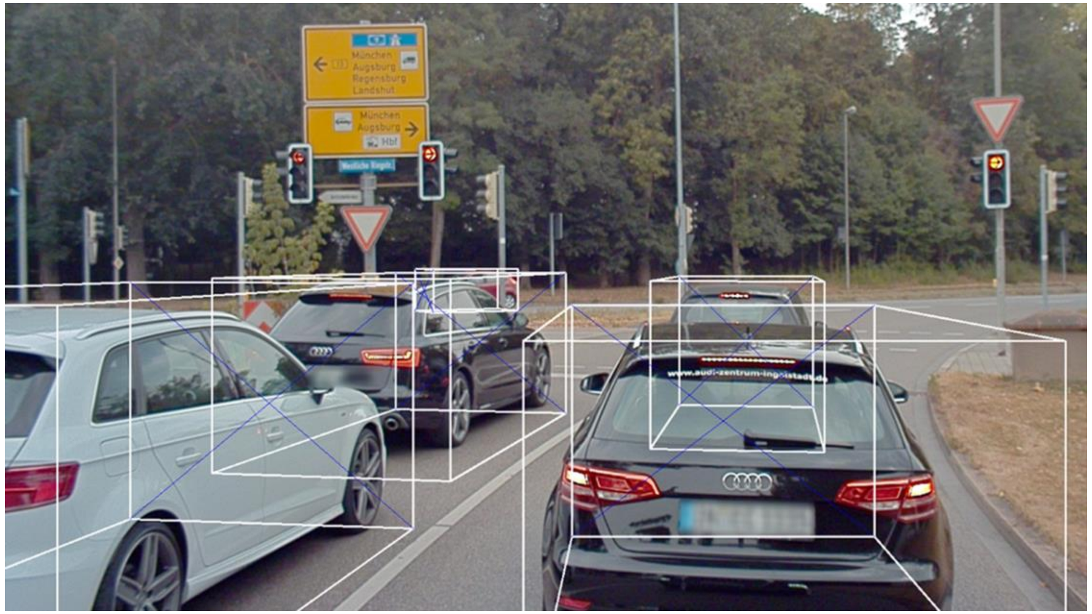
nuScenes

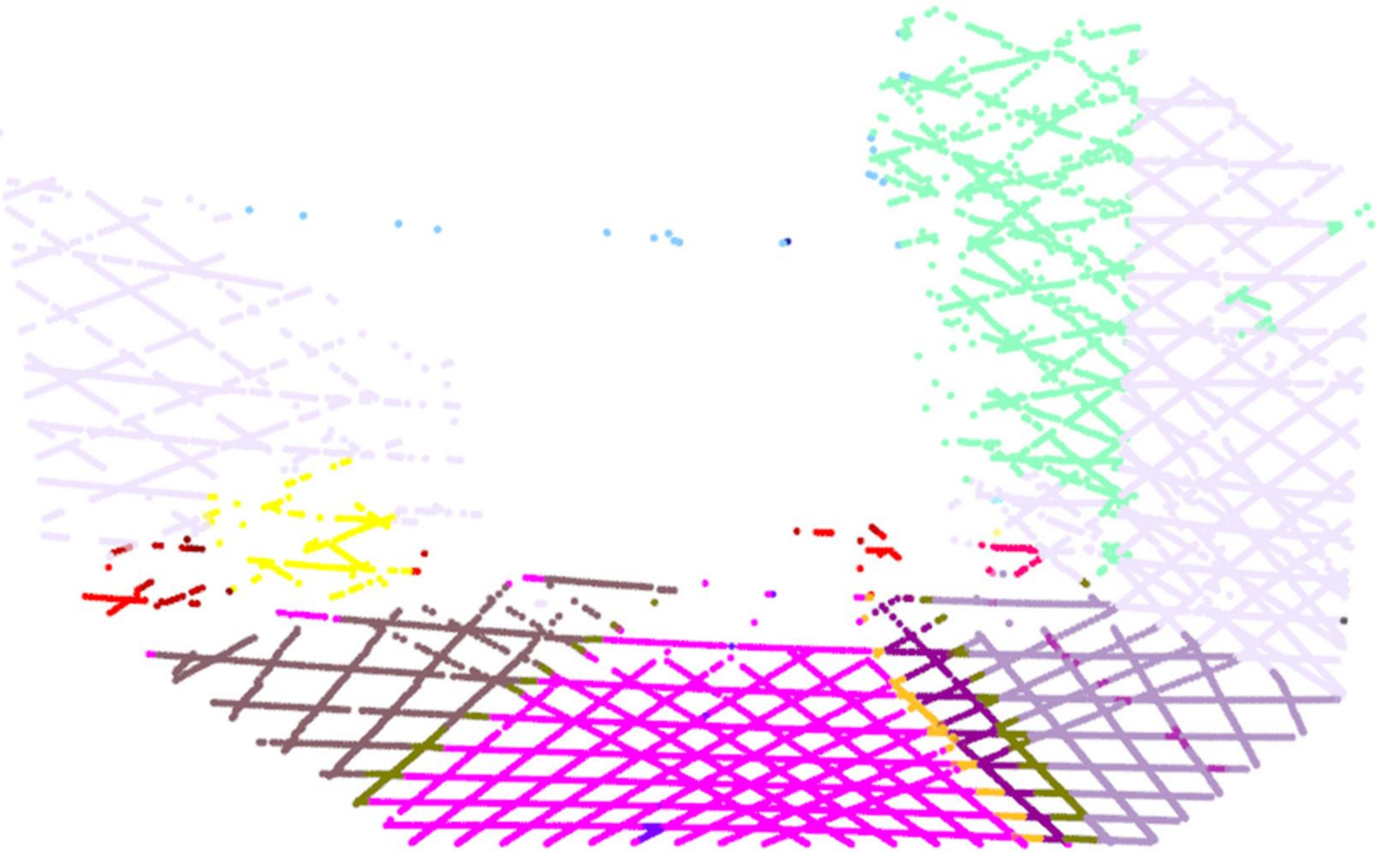
APTIV



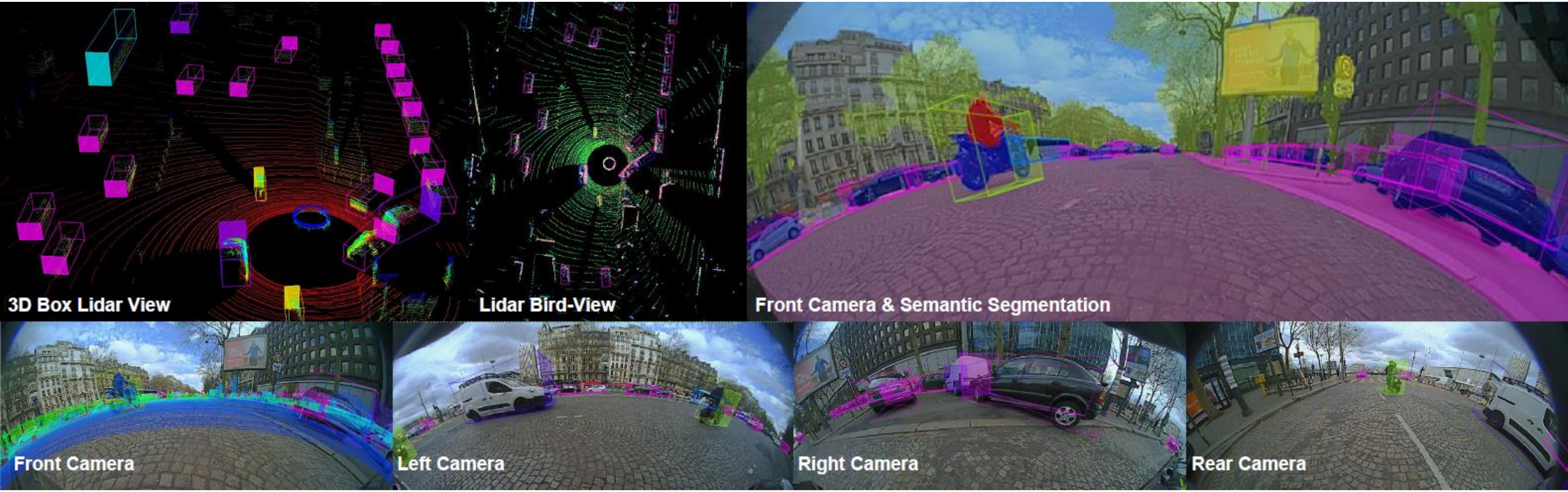
A2D2

Audi





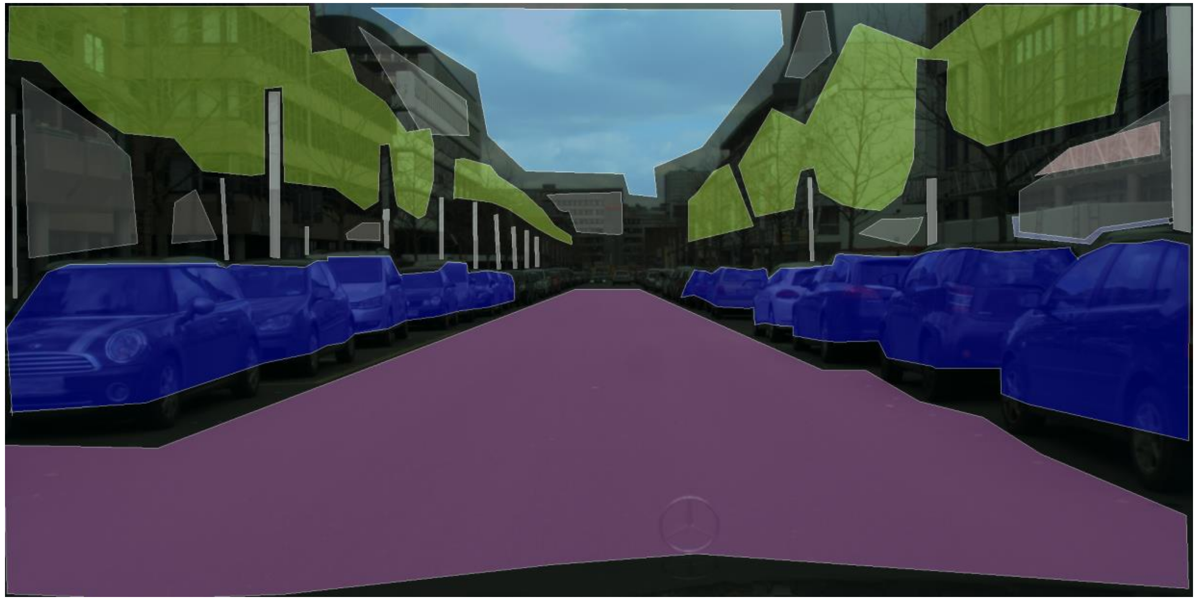
Valeo Woodscape



Annotation hell

SoA visual deep learning is fully supervised

- Data collection is not so easy (complex, costly, possibly dangerous)
- Labelling is hell (if possible)



- Doomed insufficient for in-the-wild, life-long visual understanding

Toward sustainable supervision?

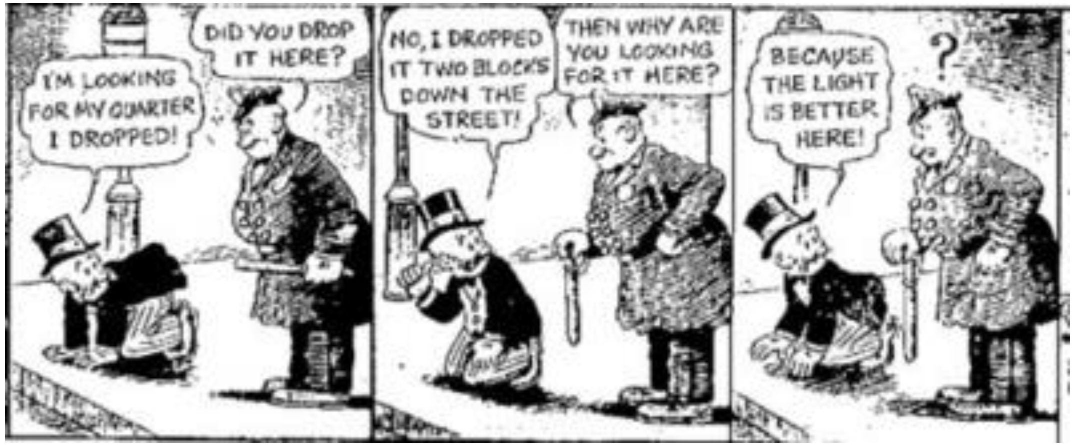
Alternatives to reduce annotation needs

- Semi-supervised learning
- Unsupervised learning
- Weakly supervised learning
- Zero-shot and few-shot learning
- Transfer learning
- Domain adaptation
- Learning from synthetic data
- Self-supervised learning
- Active learning
- Incremental learning
- Online learning

Transfer and adaptation

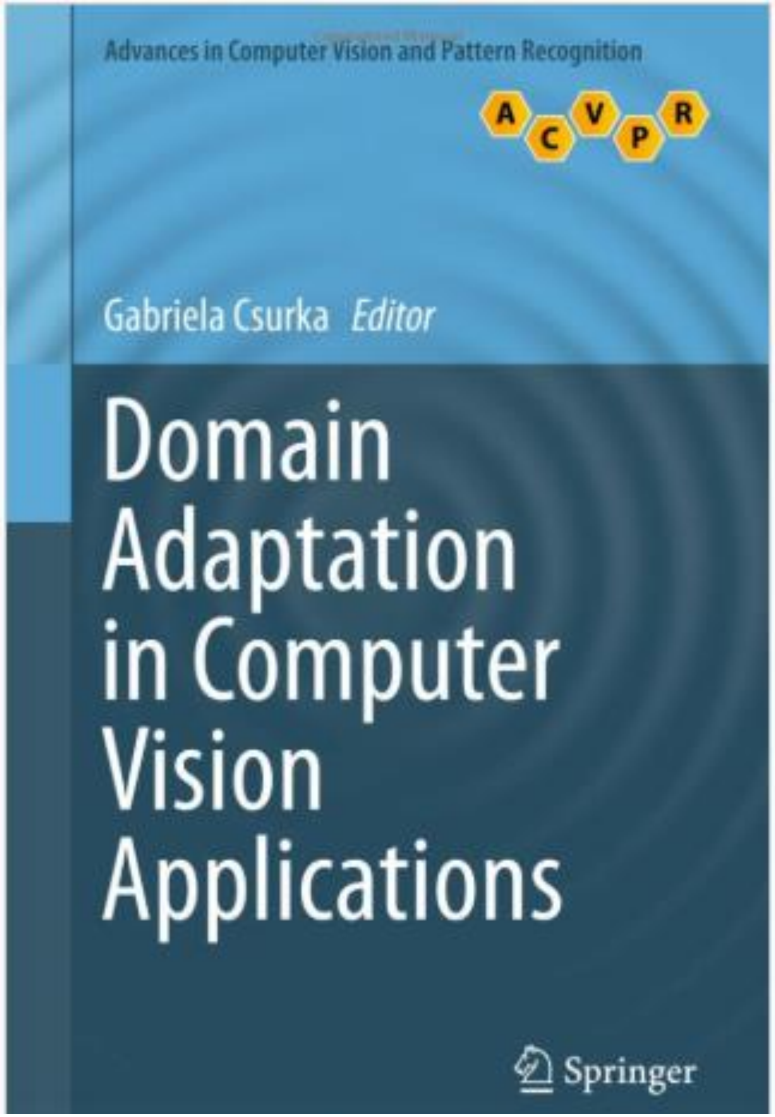
- Learn one task, conduct another one
- Learn on one distribution, run on another one = Domain Adaptation

Street light effect (a.k.a. drunkard's search)?



Not quite….

Domain adaptation in vision



Domain gap

Different, though *related* input data distributions

Source domain → Target domain

- Different weather, light, location, sensor’s spec/setup
- Synthetic vs. real

Domain gap

Different, though *related* input data distributions

Source domain → Target domain

- Different weather, light, location, sensor’s spec/setup
- Synthetic vs. real

Domain gap

Different, though *related* input data distributions

Source domain → Target domain

- Different weather, light, location, sensor's spec/setup
- Synthetic vs. real

Domain gap

Different, though *related* input data distributions

Source domain → Target domain

- Different weather, light, location, sensor's spec/setup
- Synthetic vs. real

Domain gap

Different, though related input data distributions

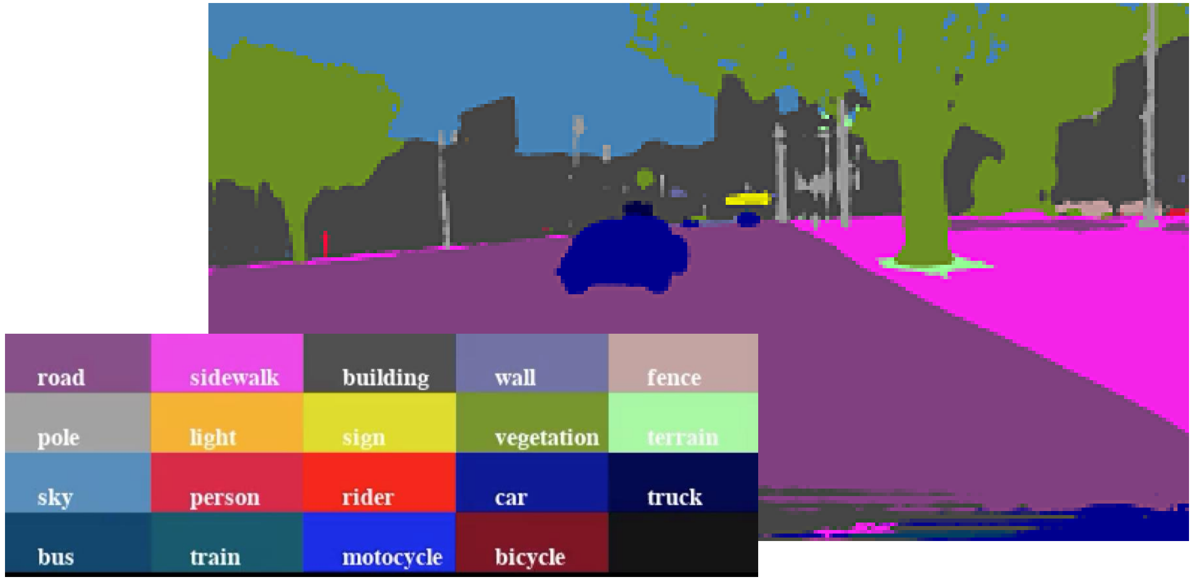
Source domain → Target domain

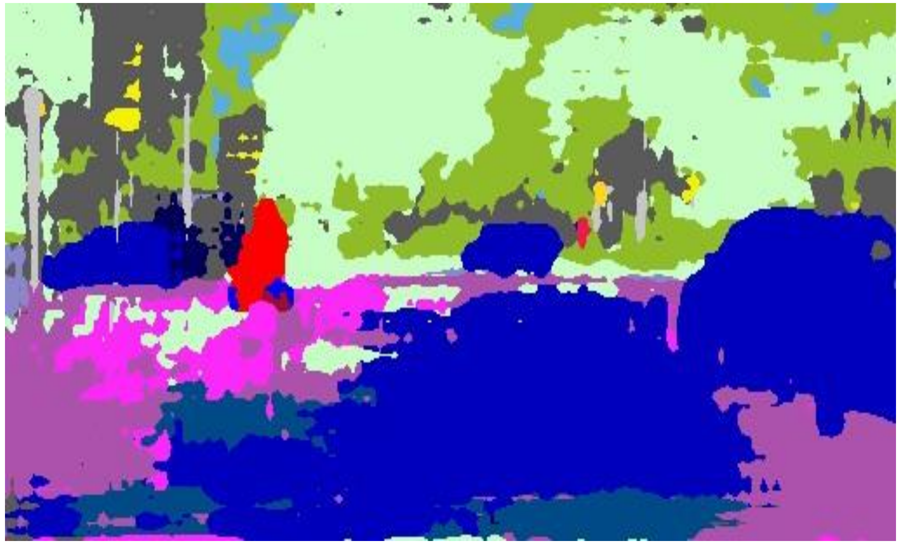
- Different weather, light, location, sensor’s spec/setup
- Synthetic vs. real

Domain gap

Different, though related input data distributions

Source domain → Target domain

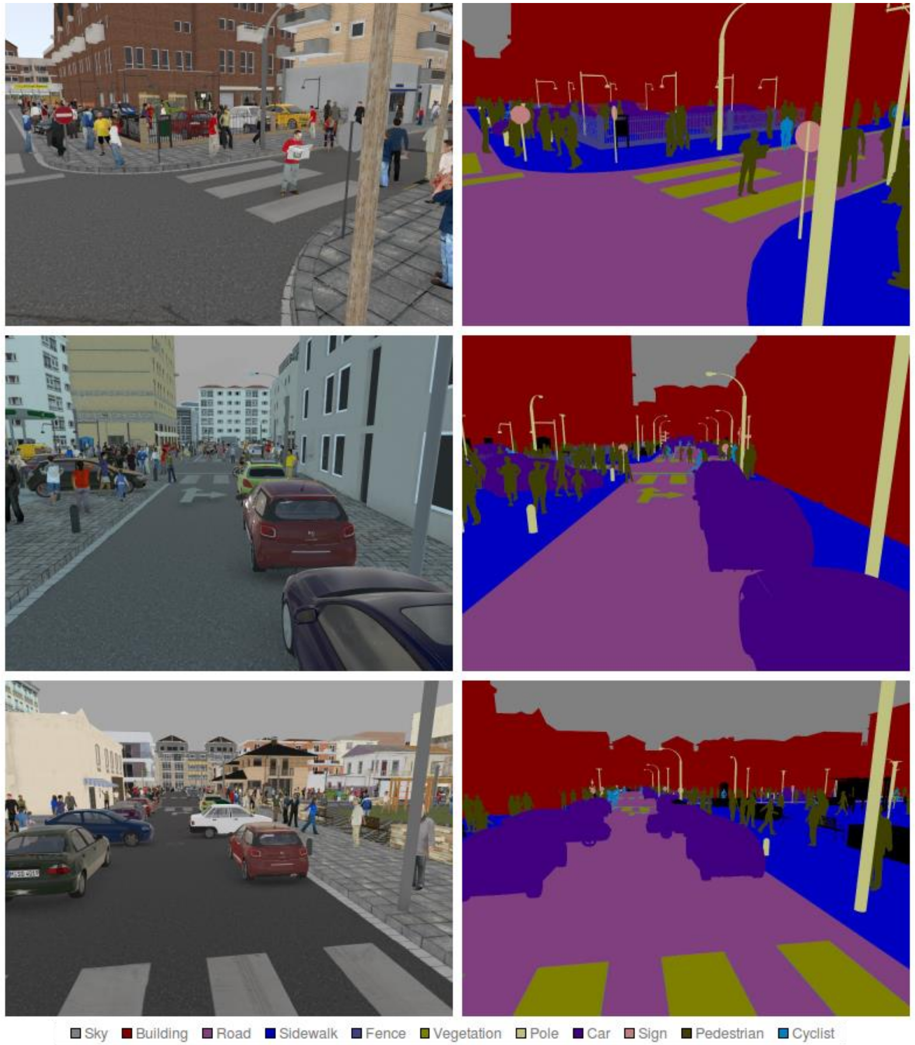




- Different weather, light, location, sensor’s spec/setup
- Synthetic vs. real

Unsupervised Domain Adaptation (UDA)

Labelled source domain data



Unlabelled target domain data

Deep learning for UDA

Distribution alignment

- Appearance, deep features, outputs

Some alignment tools

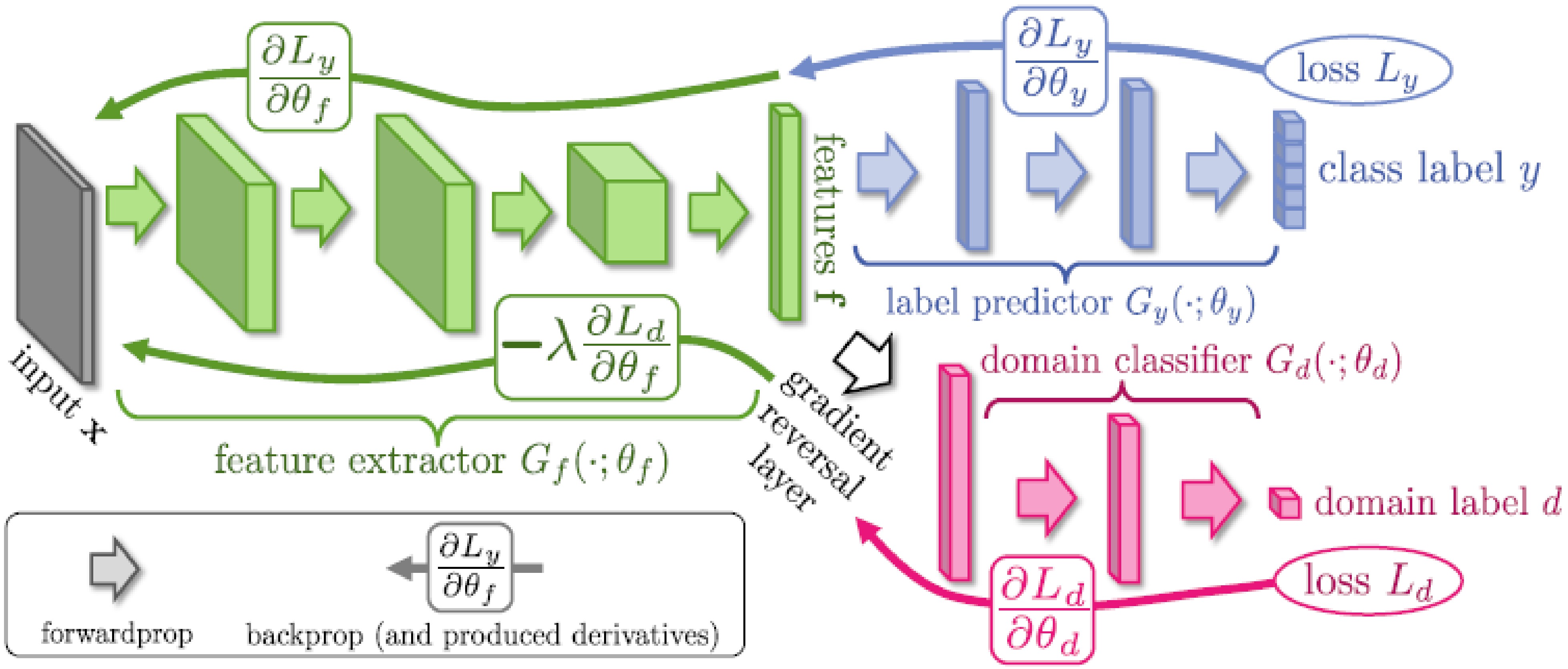
- Distribution discrepancy loss
- Optimal transport
- Discriminative adversarial loss
- Generative adversarial models

Self-training

- Curriculum learning
- Pseudo-label from confident prediction on target data

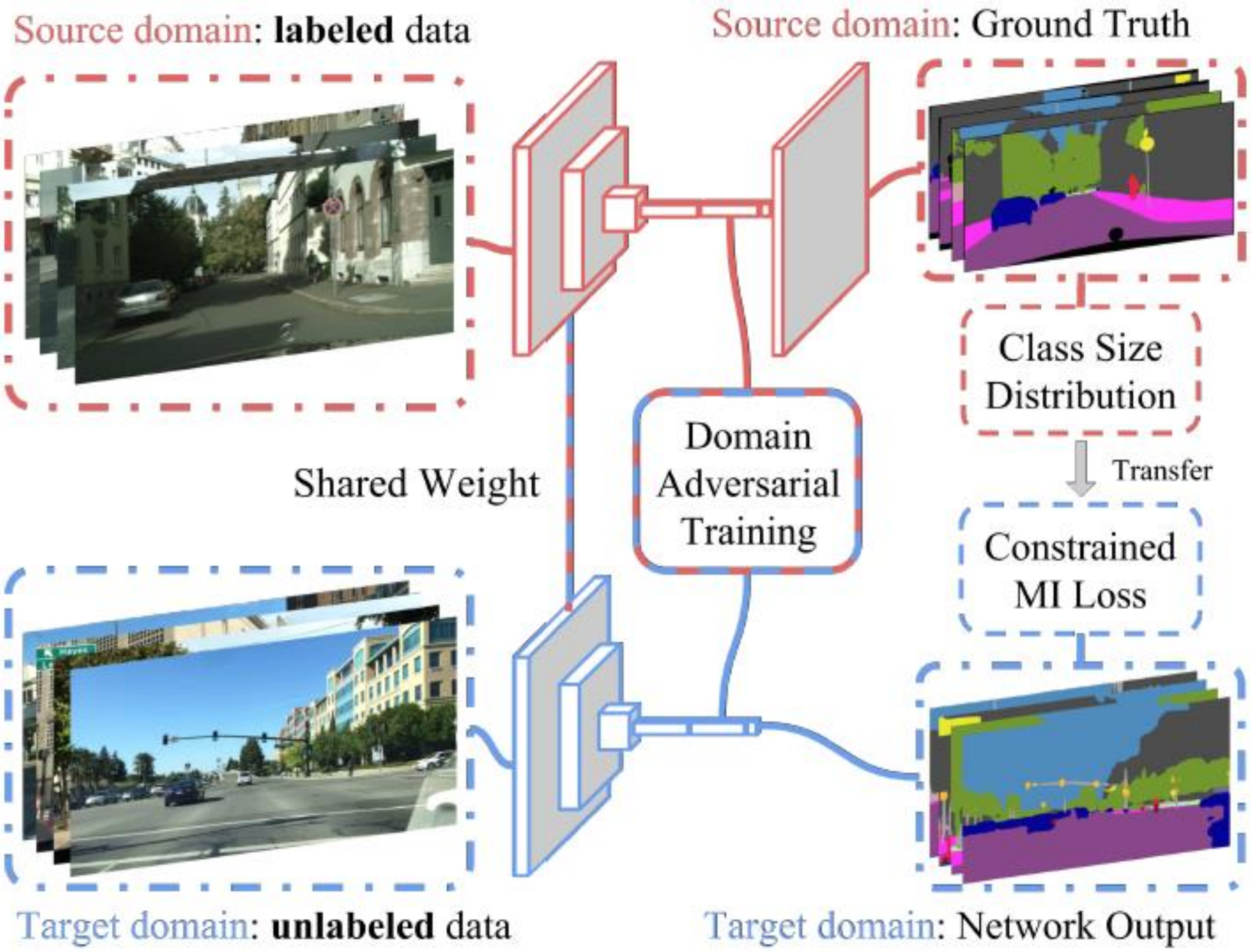
Adversarial gradient reversal

[Ganin ICML'15]



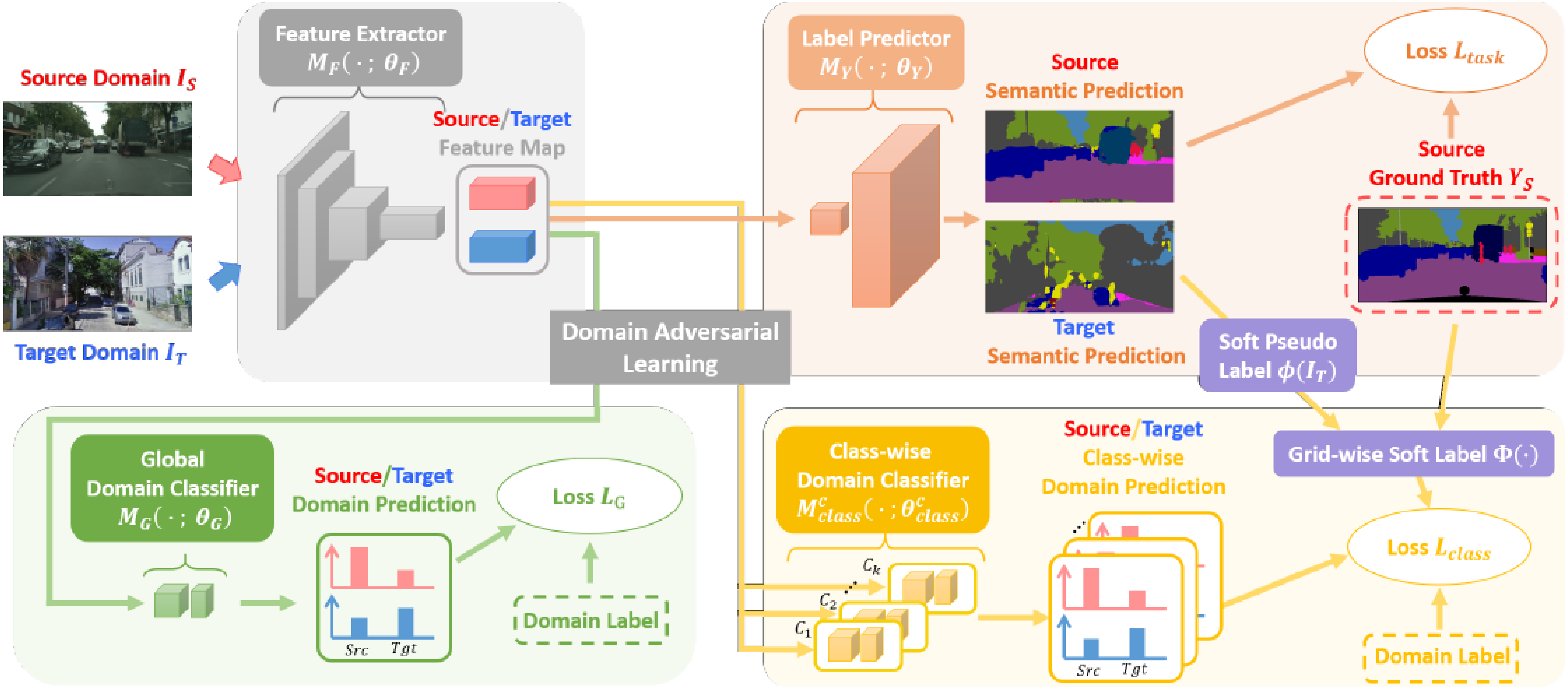
Adversarial feature alignment

[Hoffmann 2016]



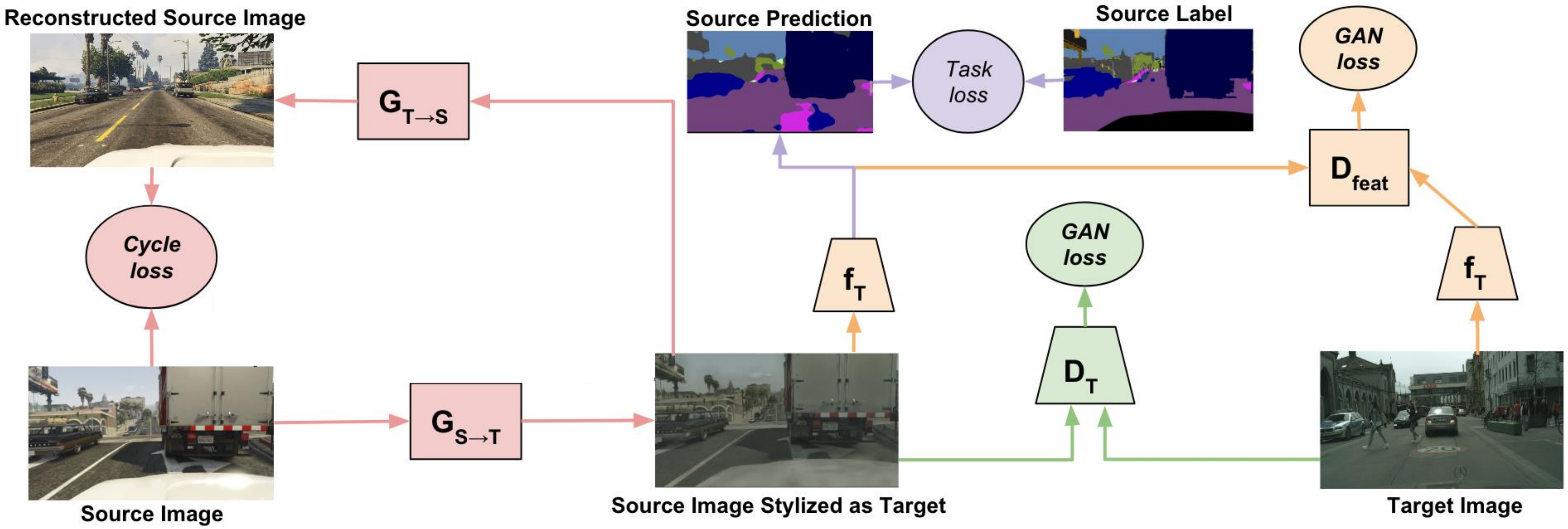
Adversarial feature alignment

[Chen ICCV'17]



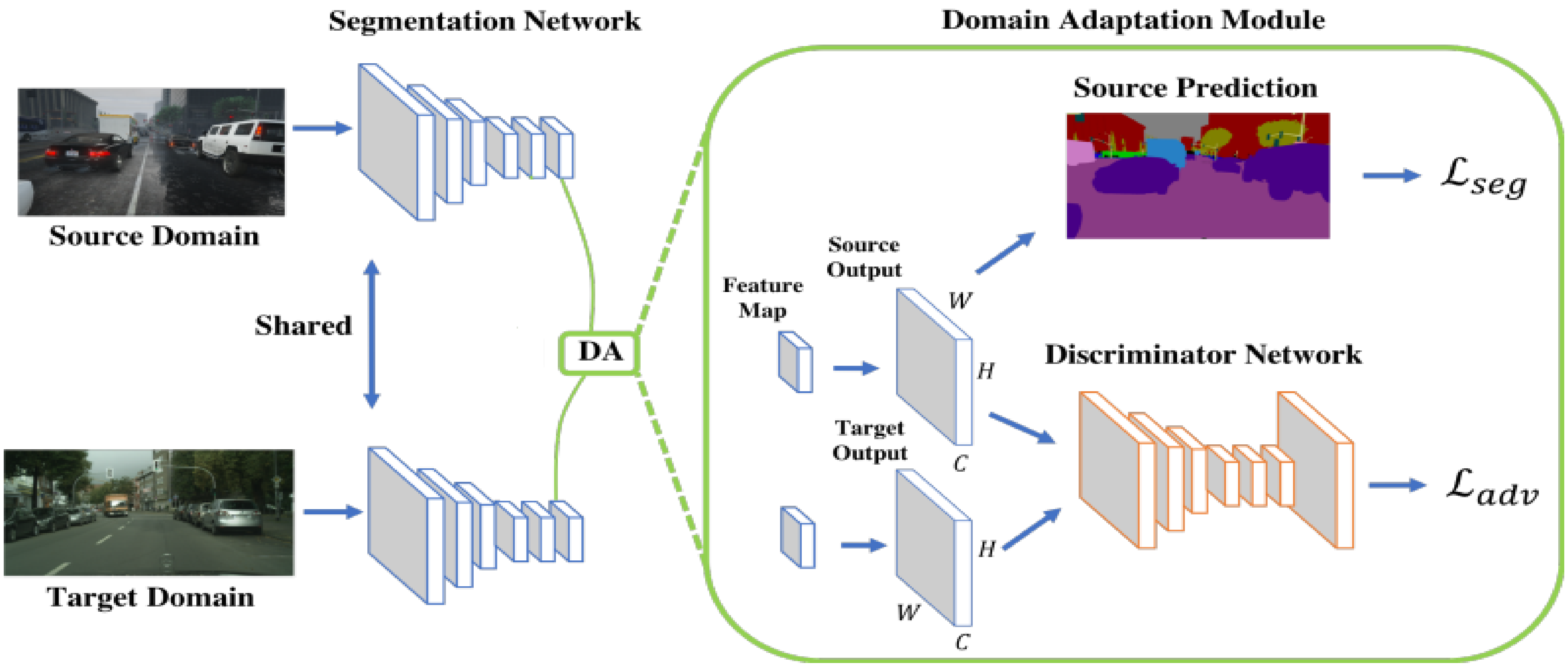
Cycle-Consistent Adversarial Domain Adaptation

CyCADA [Hoffman ICML'18]



Adversarial output alignment

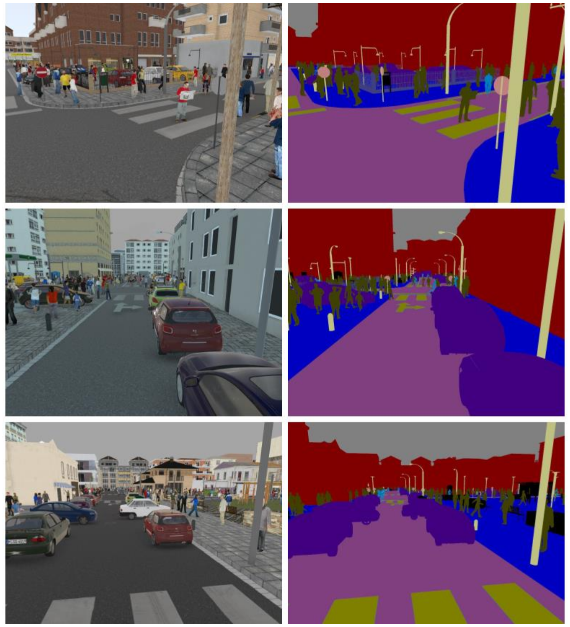
[Tsai CVPR'18]

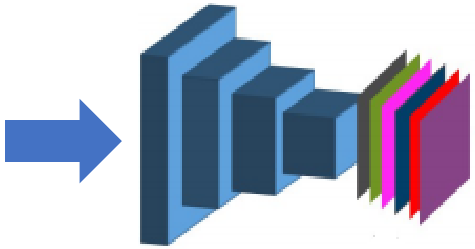


AdvEnt: Entropy-based alignment [Vu CVPR'19]

TRAIN

Source labelled data

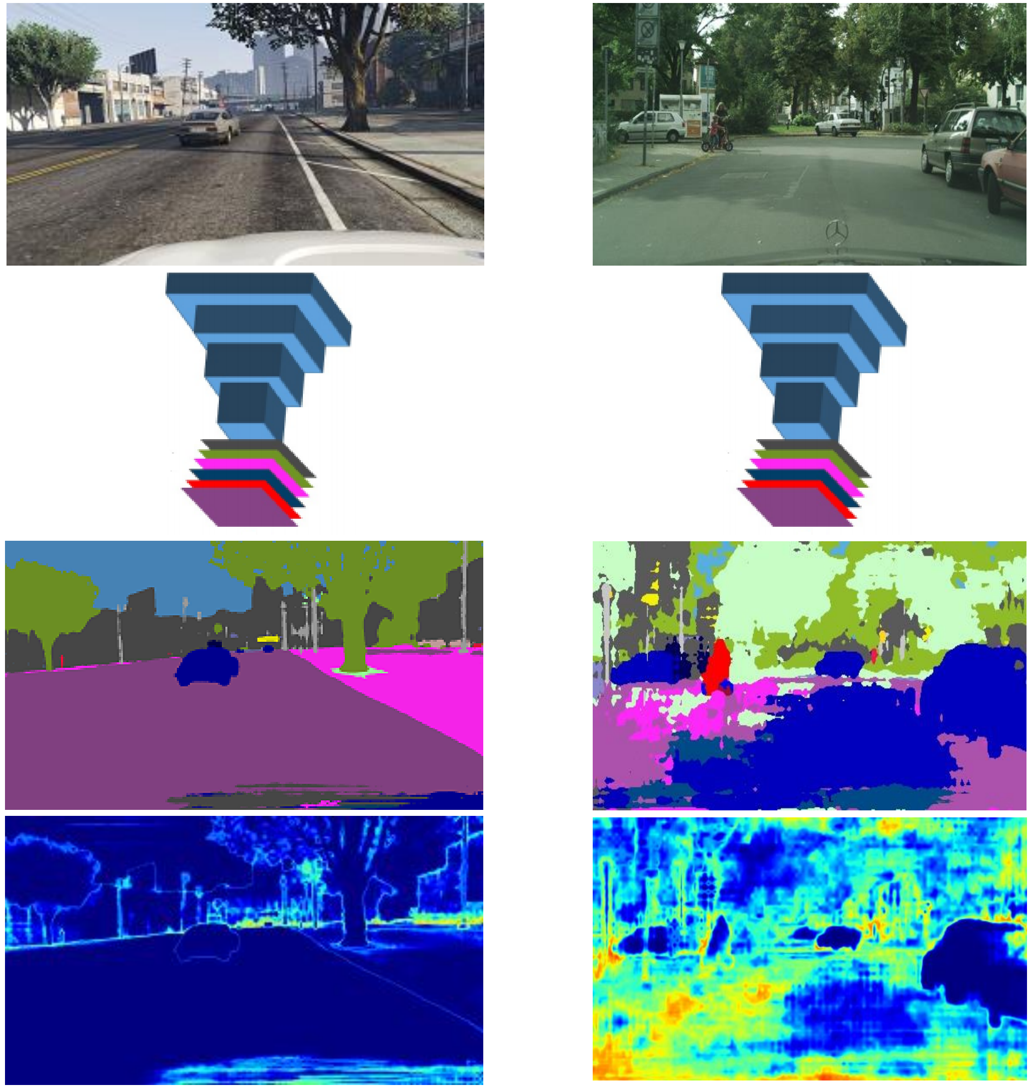




learned
segmentation
model

TEST

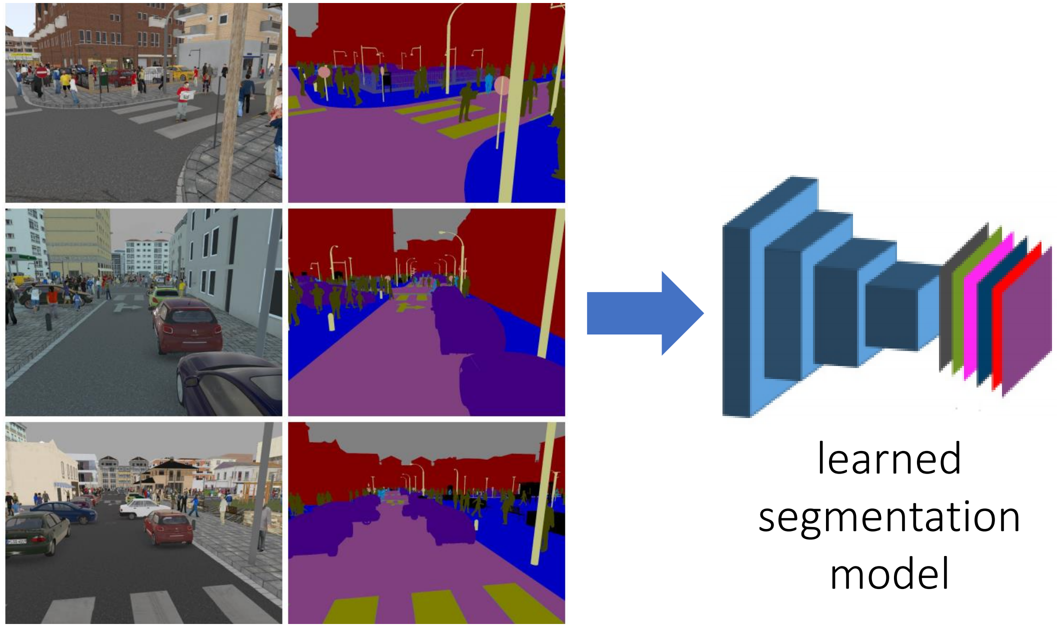
Source | Target



AdvEnt: Entropy-based alignment [Vu CVPR'19]

TRAIN

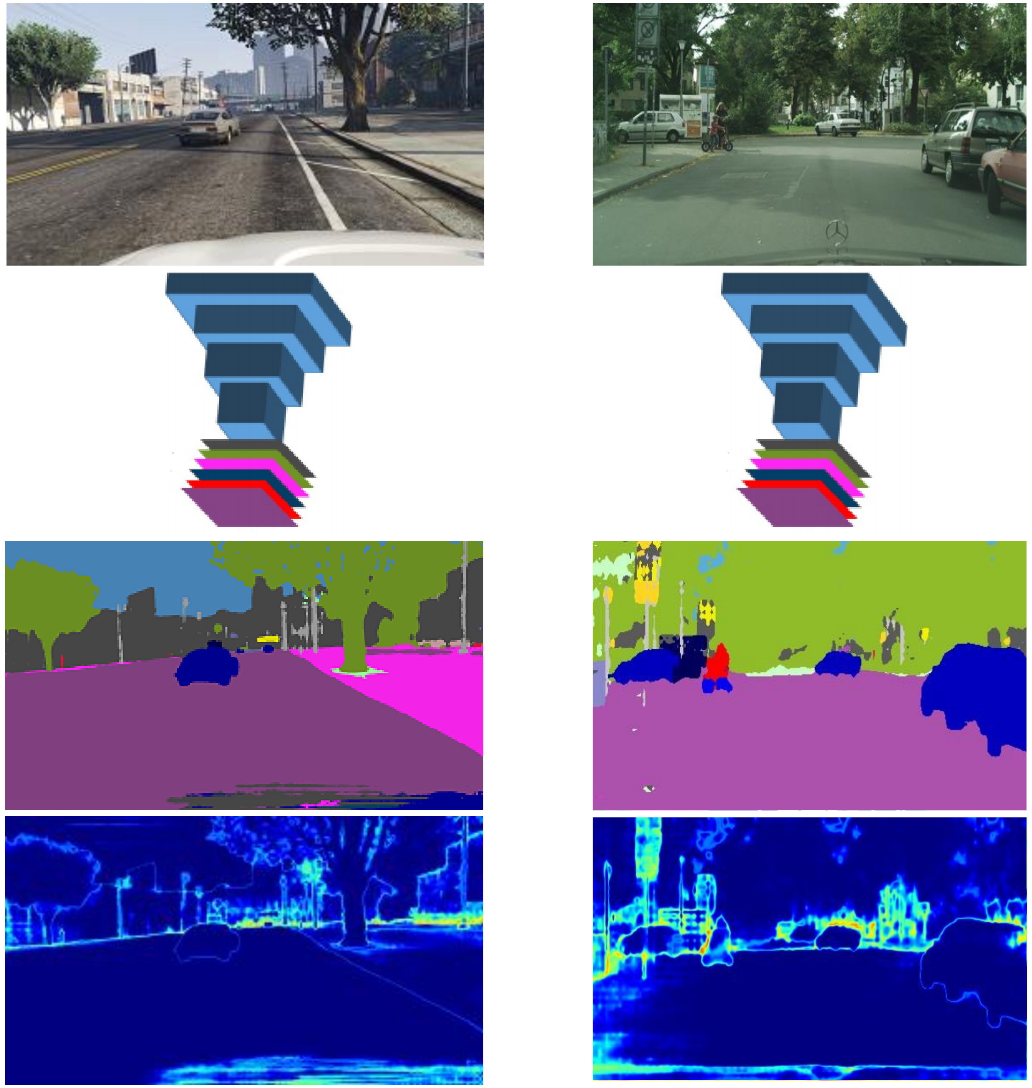
Source labelled data



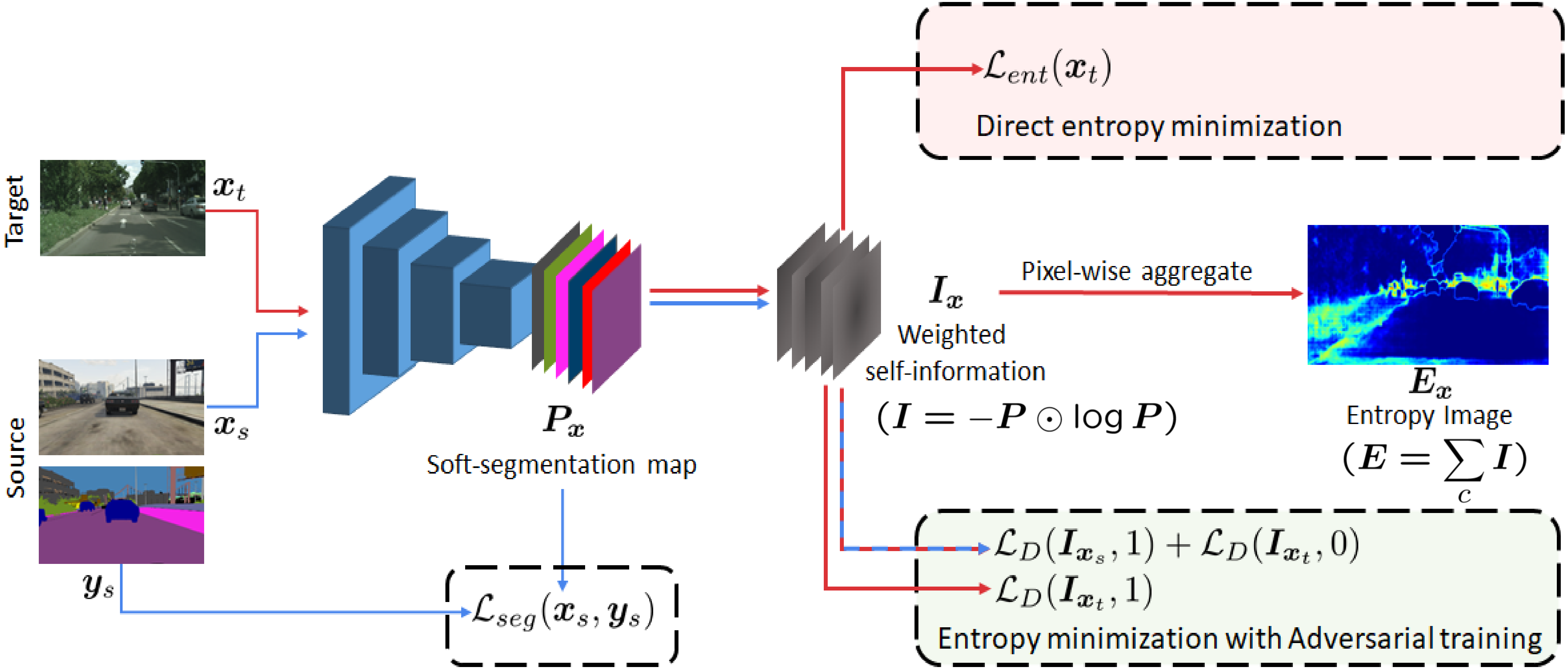
TEST

Source

Target

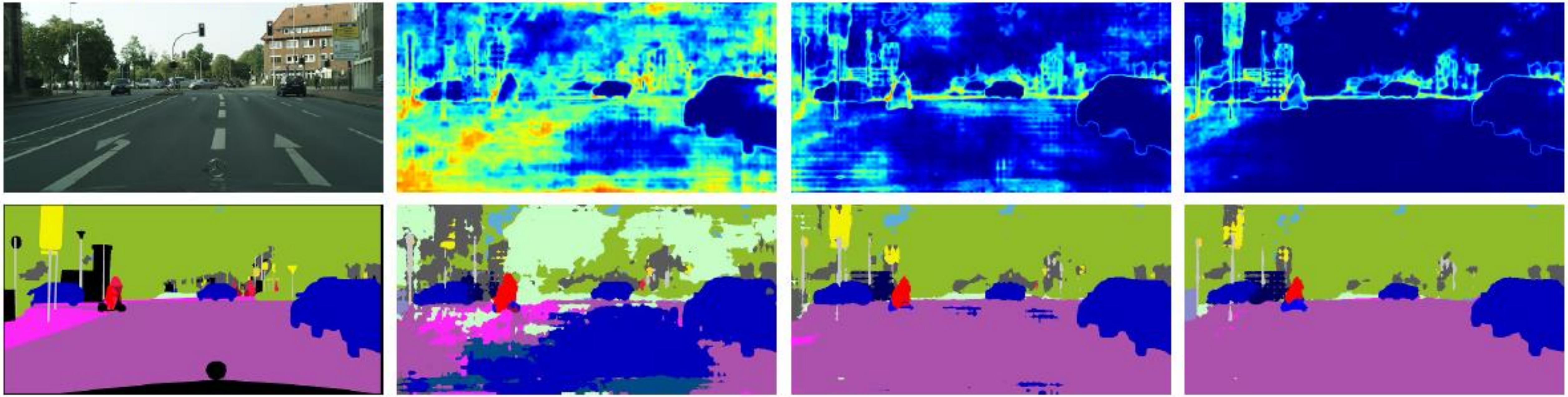


Proposed method

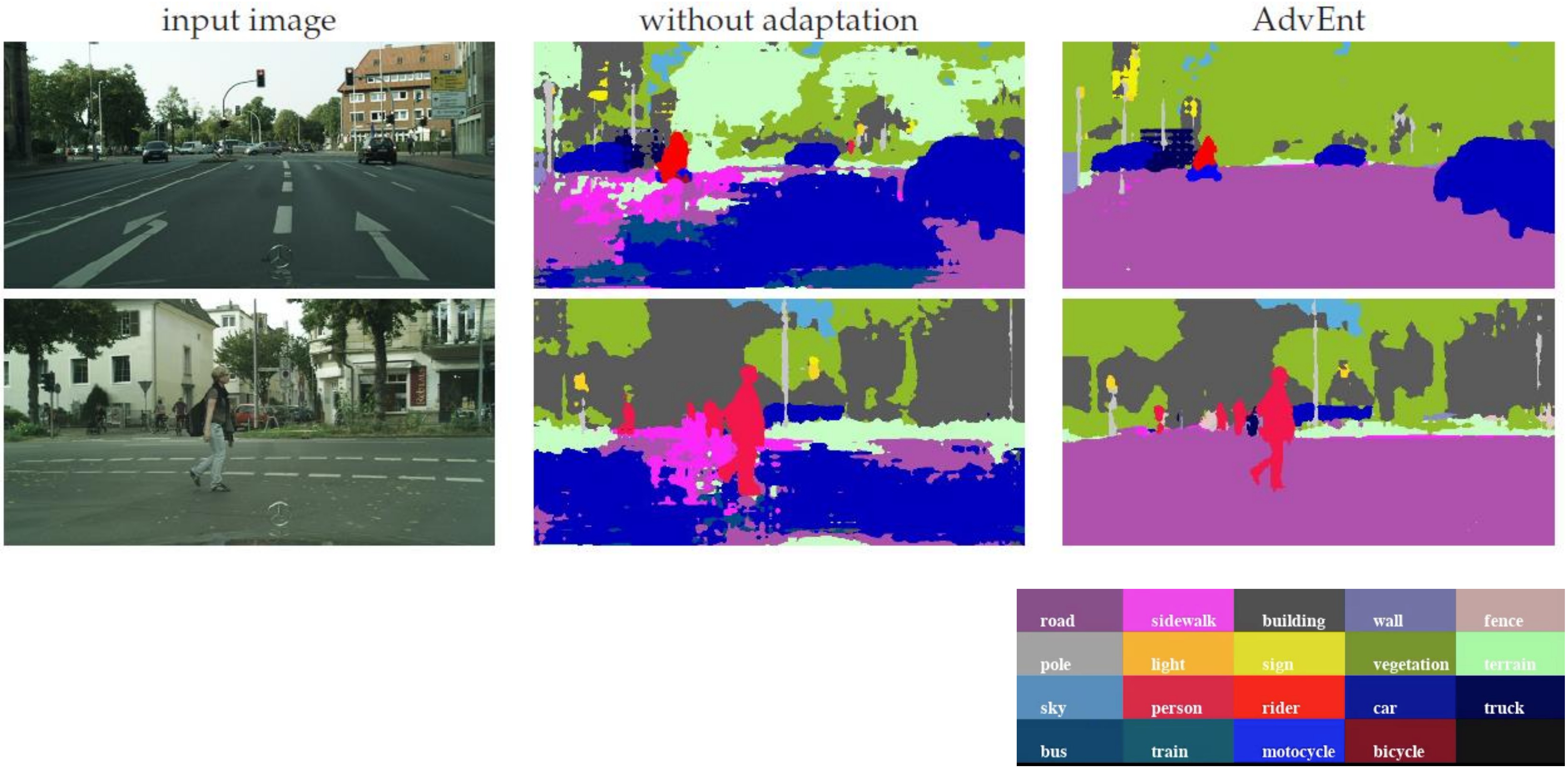


Qualitative results

(a) Input image + GT (b) Without adaptation (c) MinEnt (d) AdvEnt



Qualitative results



Extension to object detection

Clear-to-foggy-weather adaptation

- Detector: SSD-300
- Data: Cityscapes>Cityscapes-Foggy (synthetic depth-aware fog)

Privileged Information (PI) for UDA

Learning using 'Privileged Information' (LUPI)

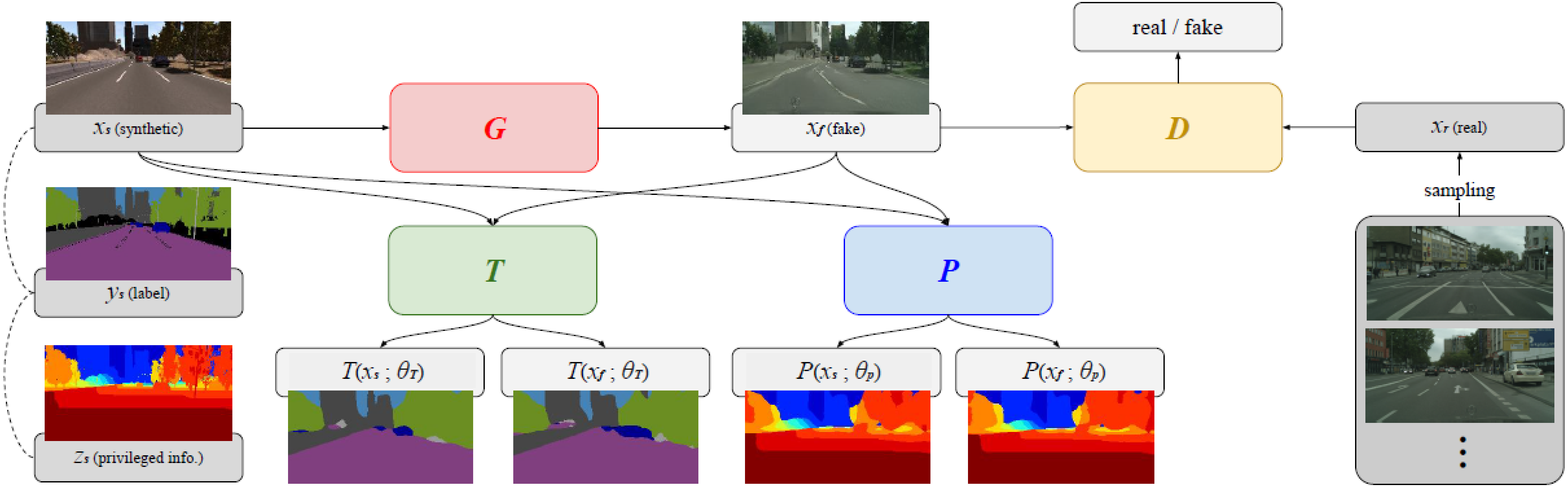
- Vapnik and Vashist 2009
- Leverage additional information at training time

In sim2real UDA

- PI comes for free on source domain, e.g. dense depth map
- Set up auxiliary task at train time (→ multi-task learning – MTL)
- Get better, domain-agnostic features
- TRI's "SPIGAN" (ICLR'19) and Vlaeo's "DADA"(ICCV19)

SPIGAN

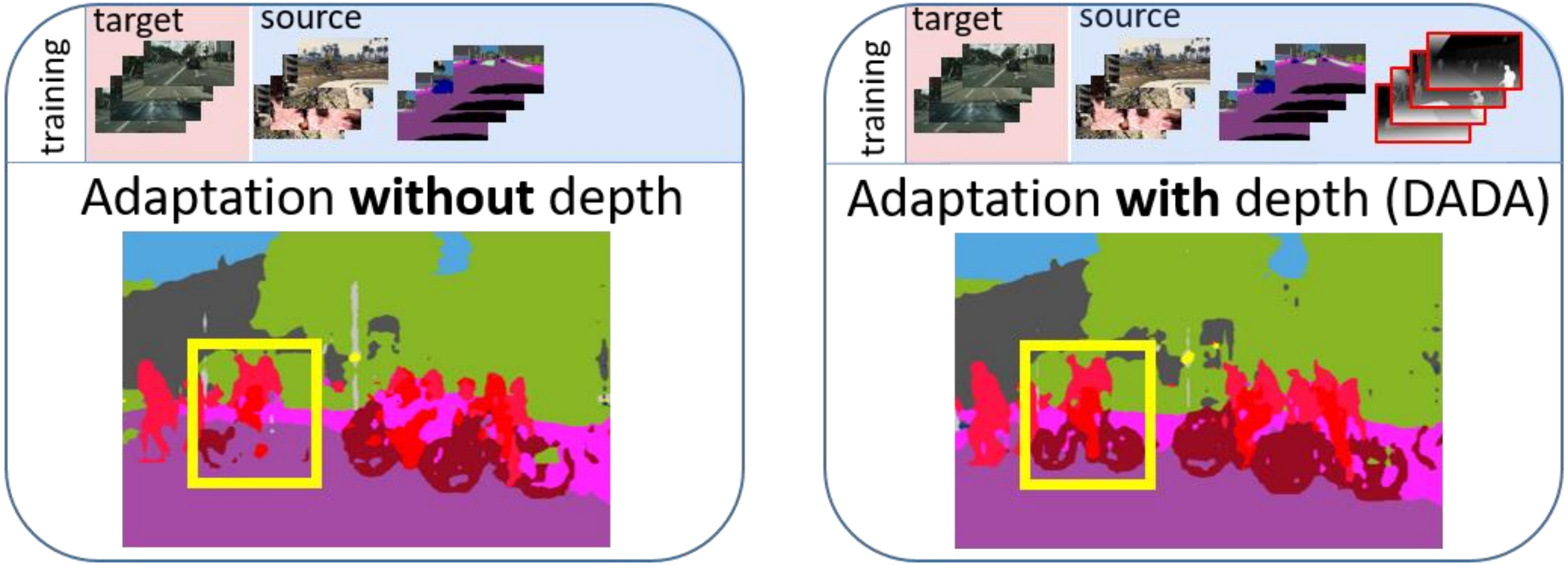
[Lee ICLR'19]



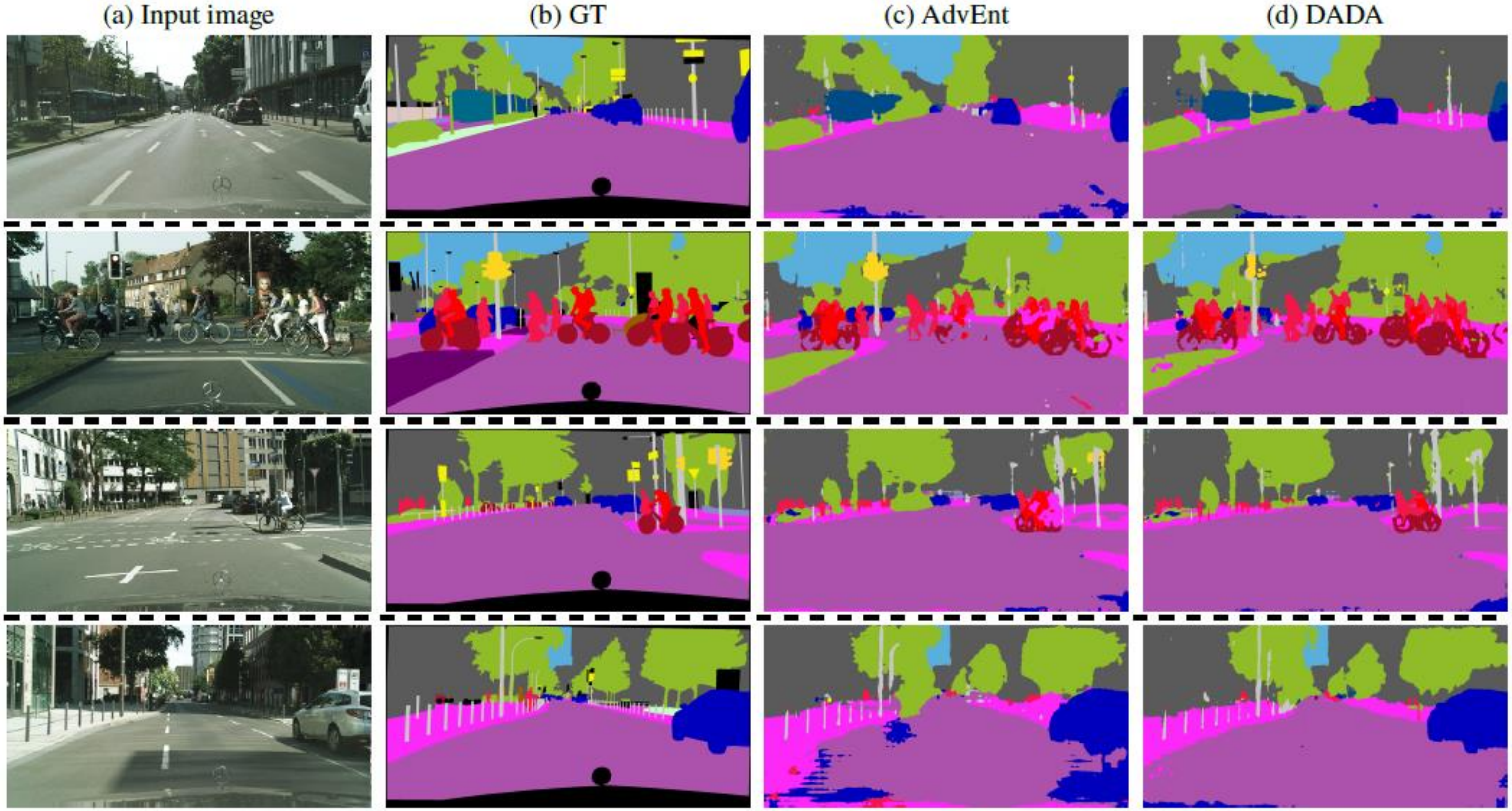
Depth-Aware Domain Adaptation (DADA)

[Vu ICCV’19]

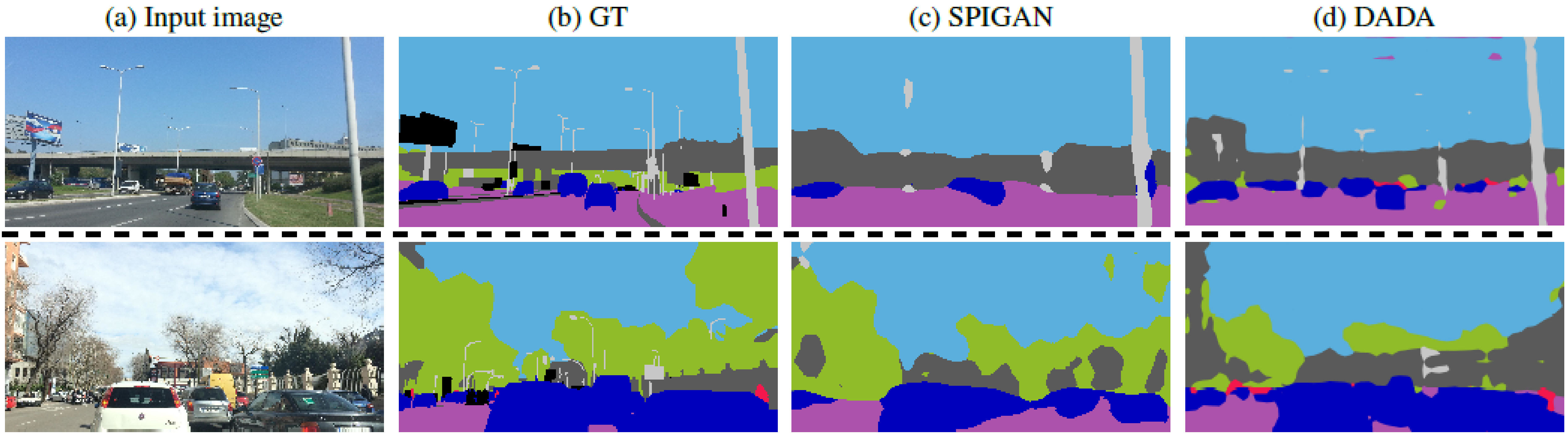
RGB image



Qualitative results

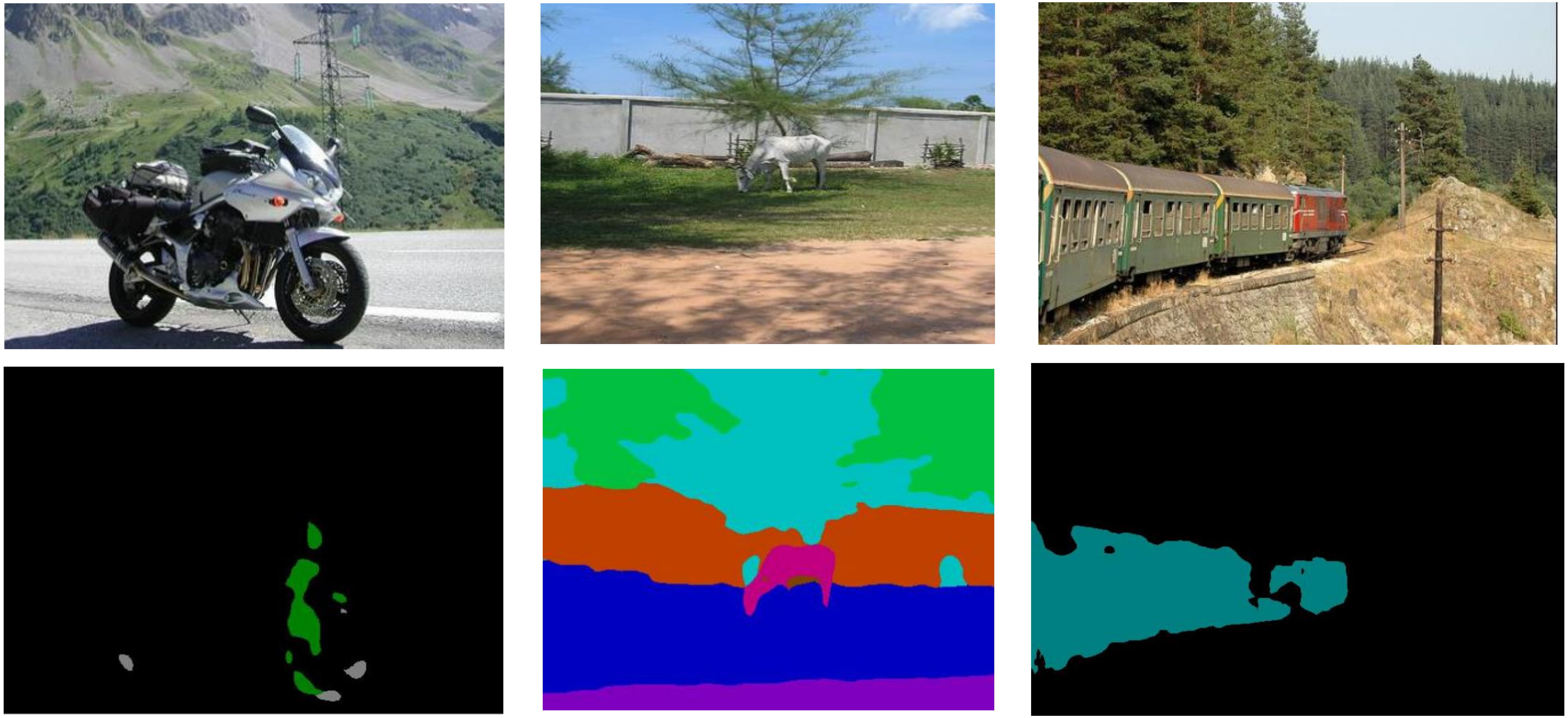


Qualitative results



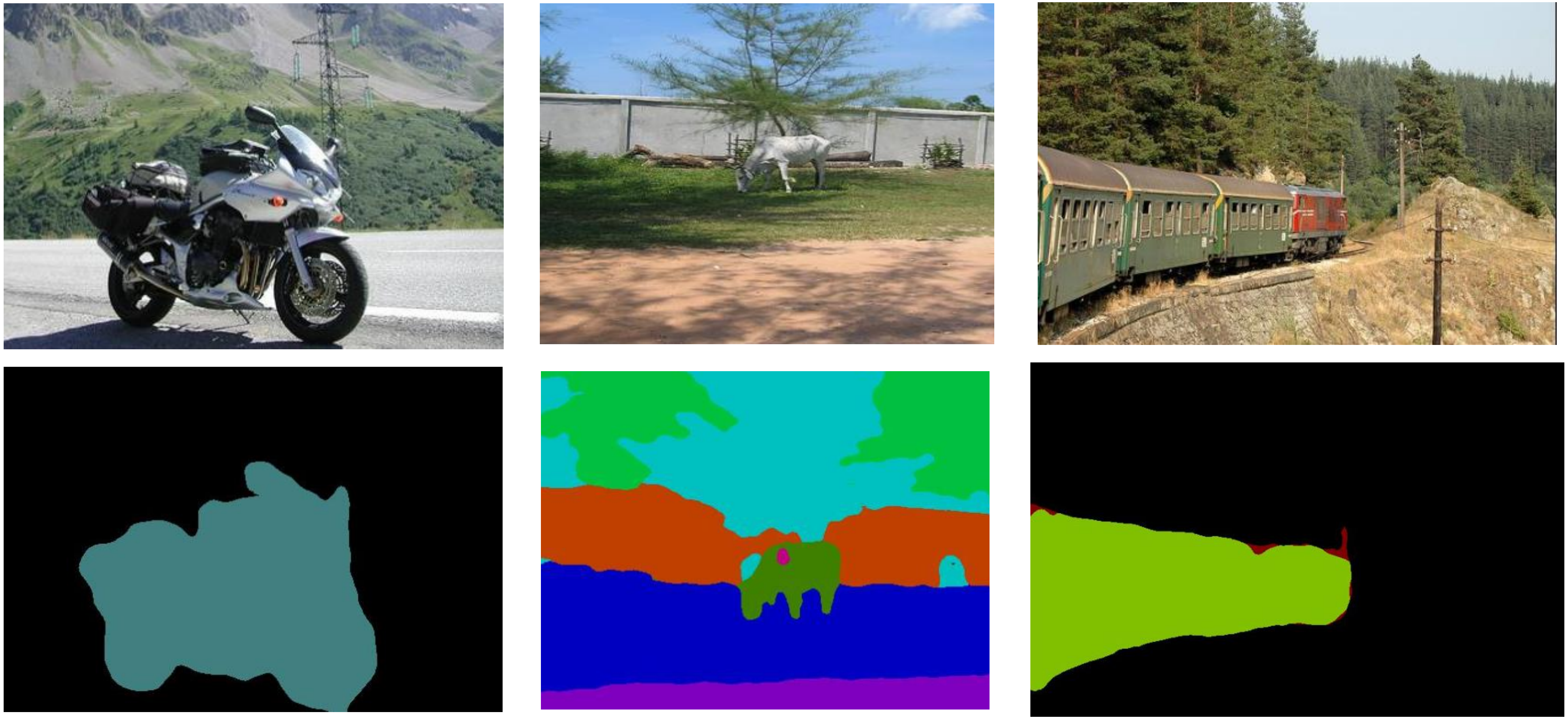
Zero-Shot Semantic Segmentation (ZS3)

[Bucher NeurIPS 2019]



Zero-Shot Semantic Segmentation (ZS3)

[Bucher NeurIPS 2019]



Take home messages and overlook

Autonomous vehicles (and robots)

- Fundamental ML challenges toward major impact
- From generalization to certified performance, on a budget

Toward sustainable supervision

- Lessen pressing need on unpractical large-scale full supervision
- Improve and hybridize low-supervision approaches
- Reduce sim2real gap
- Make RL a reality
- Leverage world knowledge