UNIVERSITY OF ALBERTA

Foreground-focused domain adaptation for object detection

Yuchen Yang, Nilanjan Ray

Department of Computing Science, University of Alberta

UNIVERSITY OF ALBERTA

Introduction

- Pervasive ‘domain gap’ in real-world applications.

- Rainy weather
- Foggy weather
- Snowy weather
- Sandstorm weather

- TypeB Camera
- TypeC Camera
- HDR Camera
- Thermal Camera

- Detection
- Clear weather
- Urban scene
- TypeA Camera
- Daytime

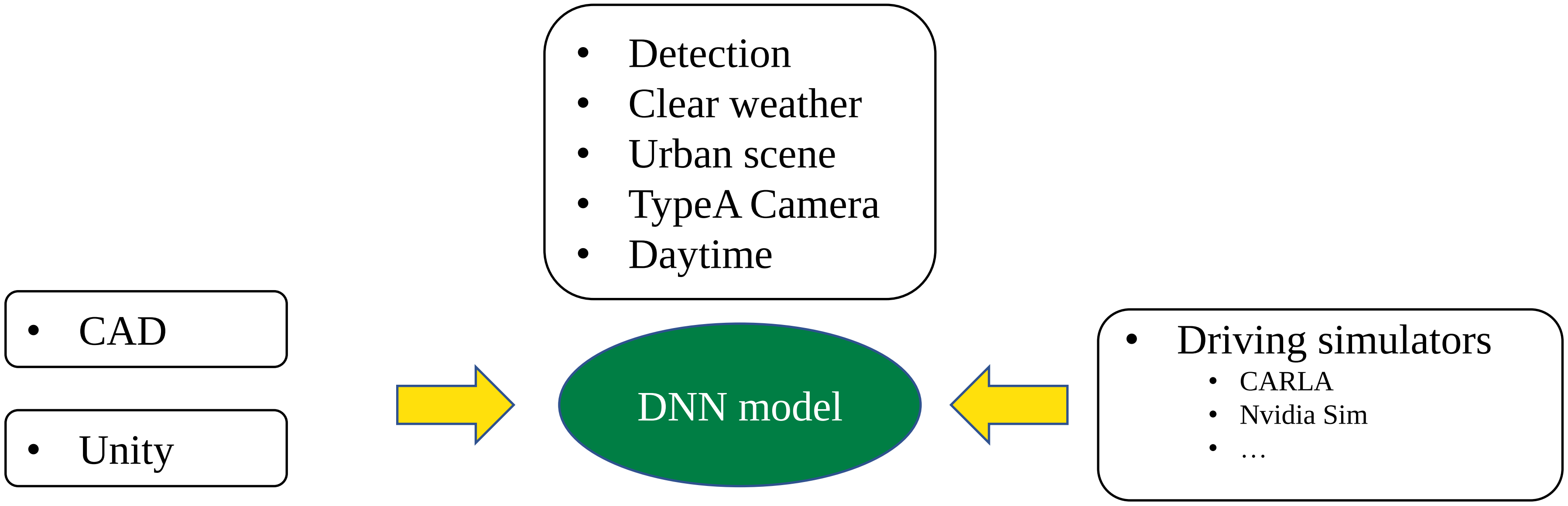
DNN model

- Highway
- Rural scene
- Race track
- Construction site

- Dusk time
- Night time

Introduction

- Pervasive ‘domain gap’ in real-world applications.



Introduction

- Unsupervised domain adaptation (UDA) object detection:
  - A detector is trained with labeled source domain images and unlabeled target domain images. Then, it is applied to detect objects in target domain images.

Source Domain $[X_s, Y_s]$

Target Domain $[X_t]$

Apply to

Target Domain

Introduction

- Previous studies exploit the adaptation on full feature:
  - Alignment on background is likely to pose additional difficulties, due to the sophisticated layout and appearance in the background

Source domain – Clean weather

Adapt

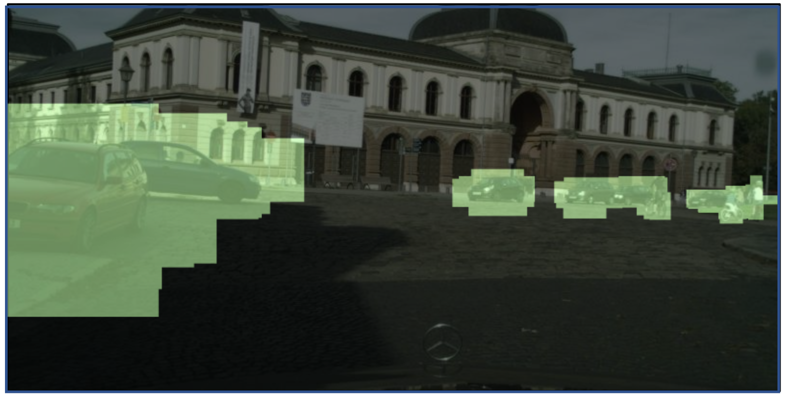
Target domain – Foggy weather

Region of adaptation

Introduction

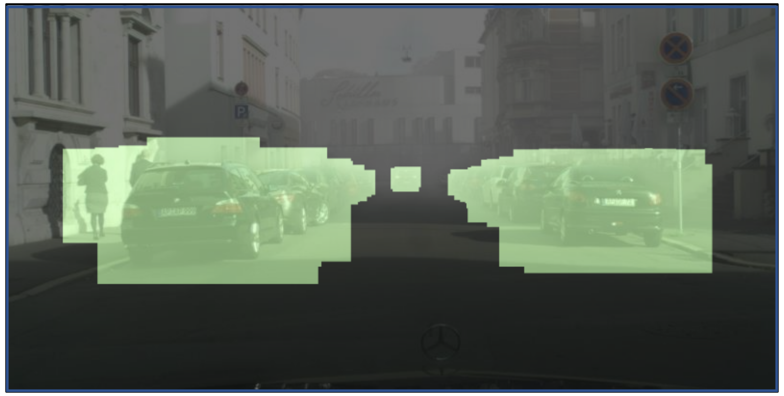
- Foreground-focused domain adaptation (FFDA):
  - We mine the loss of the domain discriminators to concentrate on the backpropagation of a foreground loss

Source domain – Clean weather



Adapt

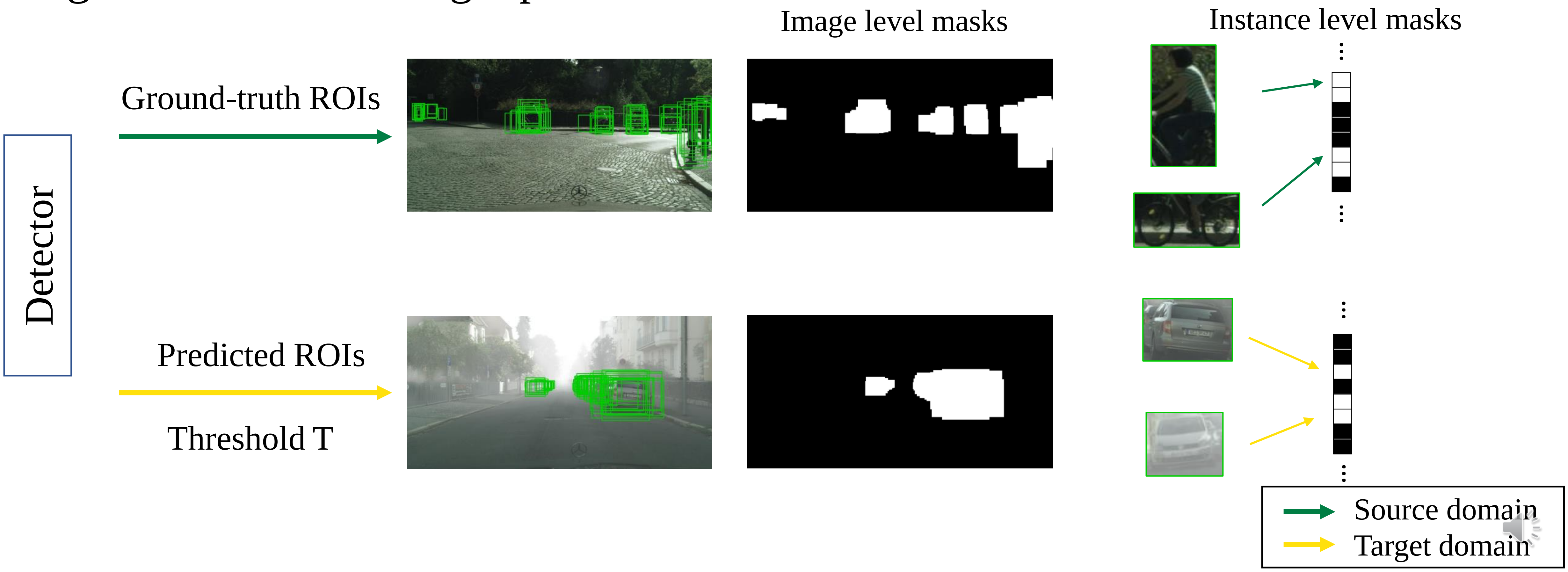
Target domain – Foggy weather



Region of adaptation

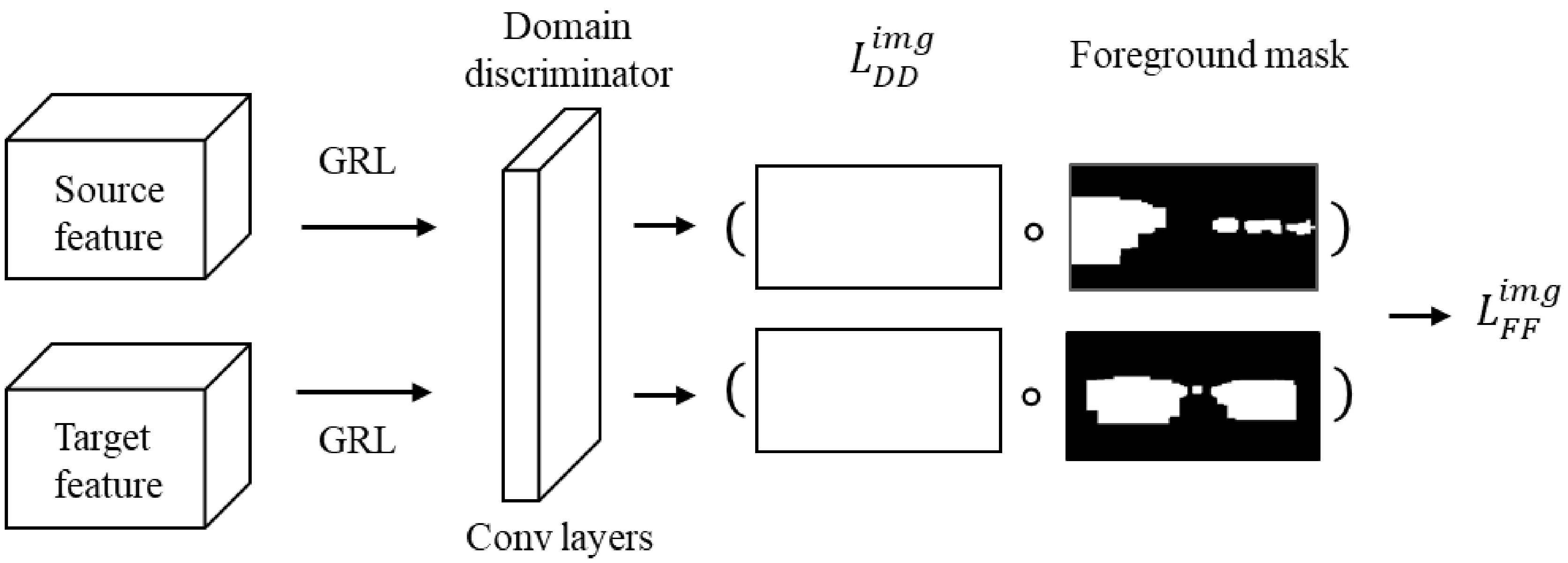
Method

- **Mining mask generation:** Mining masks are generated using source ground-truth and target predictions.



Method

- **Image level FFDA:** We mine the loss in foreground area on the loss map generated from image level domain discriminator.



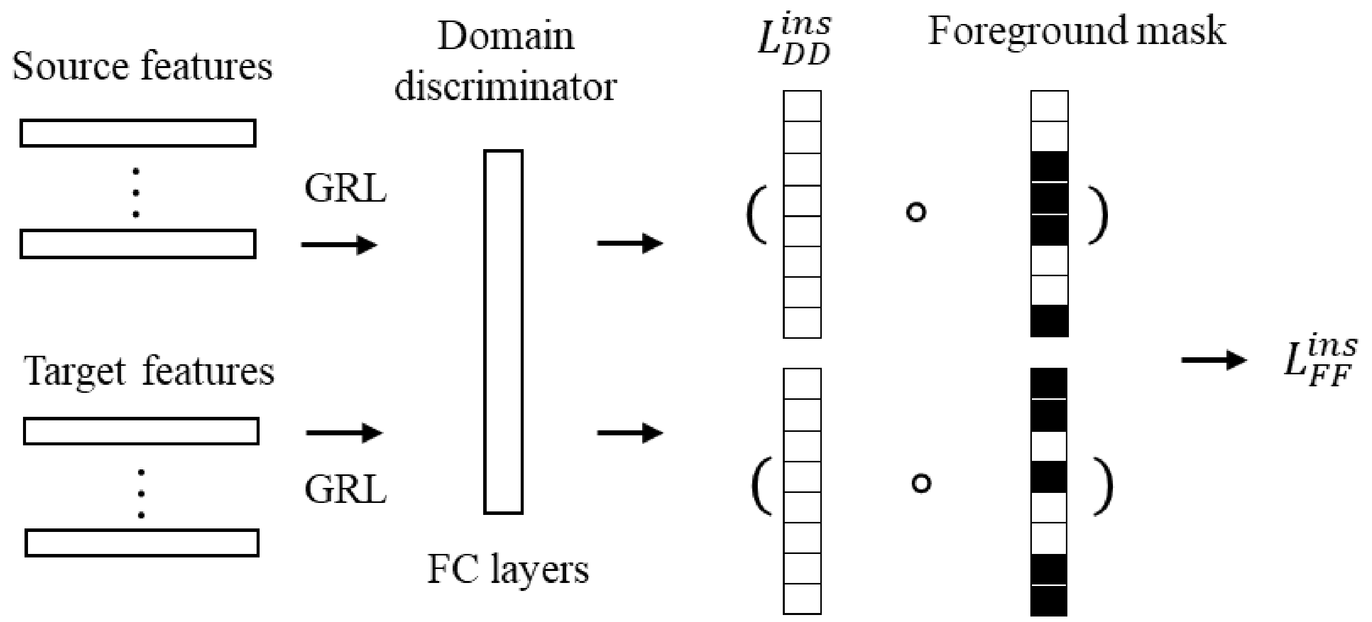
$$\min_{\theta}\max_{w} L^{img}_{FF}$$

$$L^{img}_{FF} = \frac{1}{N^{pixel}_s}\sum_{u,v} L^{img}_{DD}(u,v)_s M^{img}_s(u,v) + \frac{1}{N^{pixel}_t}\sum_{u,v} L^{img}_{DD}(u,v)_t M^{img}_t(u,v)$$

$$L^{img}_{DD}(u,v)_{s,t} = -D_i \log(p_i(u,v)) - (1 - D_i)\log(1 - p_i(u,v))$$

Method

- **Instance level FFDA:** We mine the loss of identified foreground ROI features from instance level domain discriminator.



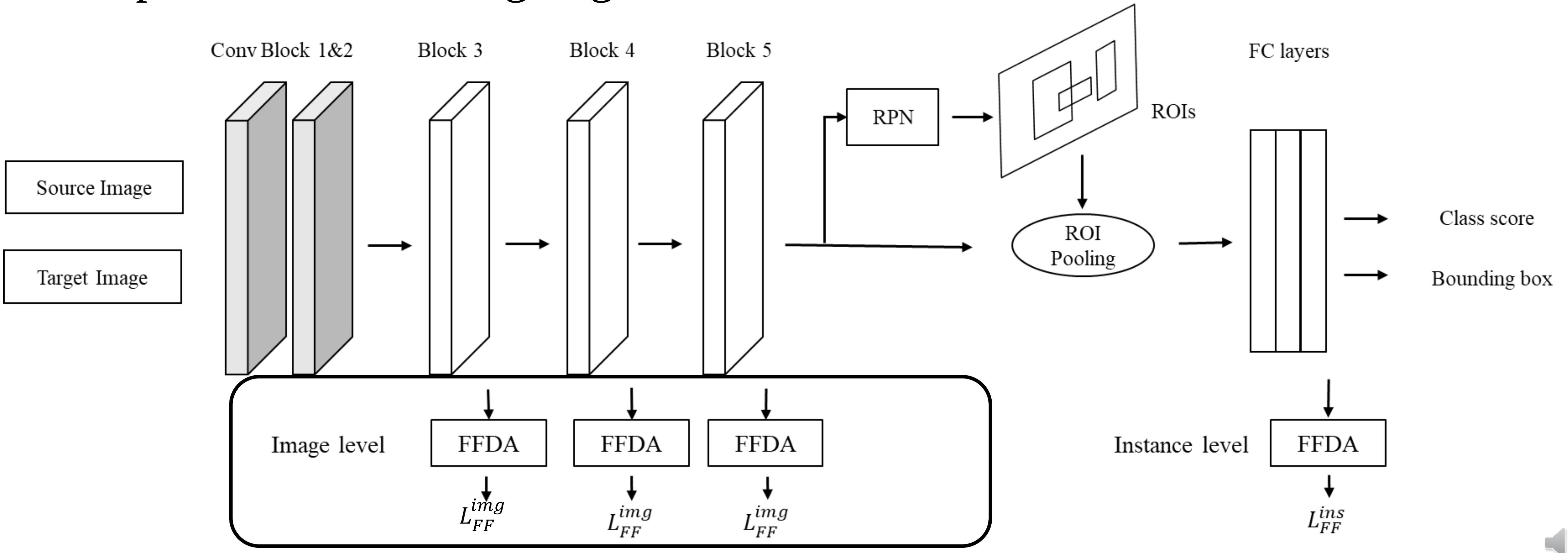
$$\min_{\theta} \max_{w} L_{FF}^{ins}.$$

$$L_{FF}^{ins} = \frac{1}{N_s^{feat}} \sum_j L_{DD}^{ins}(j)_s M_s^{ins}(j) + \frac{1}{N_t^{feat}} \sum_j L_{DD}^{ins}(j)_t M_t^{ins}(j)$$

$$L_{DD}^{img}(u,v)_{s,t} = -D_i \log(p_i(u,v)) - (1 - D_i) \log(1 - p_i(u,v))$$

Method

- **Multi-adversarial alignment:** We attach multiple image level FFDA subparts to build strong alignment on features.



Experiments

- We evaluate our method on four datasets for different scenarios in autonomous driving applications.

- Clear to Foggy weather (Cityscape -> Foggy Cityscape)
- Synthetic to real (SIM10K -> Cityscape)
- Cross camera (KITTI -> Cityscape)
- Daytime to nighttime (BDD100k daytime-> nighttime)

Experiments

- Mean average precision compared with previous SOTA and MLDA baseline.

Table 1. Adaptation from Cityscape to Foggy Cityscape

Methods	person	rider	car	truck	bus	train	mcycle	bicycle	mAP
Source trained	24.2	29.5	31.4	10.1	14.3	9.1	13.4	27.7	20.0
ART+PSA	**34.0**	46.9	**52.1**	**30.8**	43.2	29.9	34.7	**37.4**	38.6
MLDA	33.2	44.2	44.8	28.2	41.8	28.7	30.5	36.5	36.1
Ours (block4,5+ins)	33.8	45.6	50.6	25.2	46.0	31.3	**35.8**	**37.4**	38.2
Ours (block3,4,5+ins)	33.8	**48.3**	50.7	26.6	**49.2**	**39.4**	**35.8**	36.8	**40.1**
Oracle	36.2	45.8	52.7	33.4	51.5	44.0	37.8	39.0	42.6

Table 2. Adaptation from SIM10K to Cityscape and KITTI to Cityscape

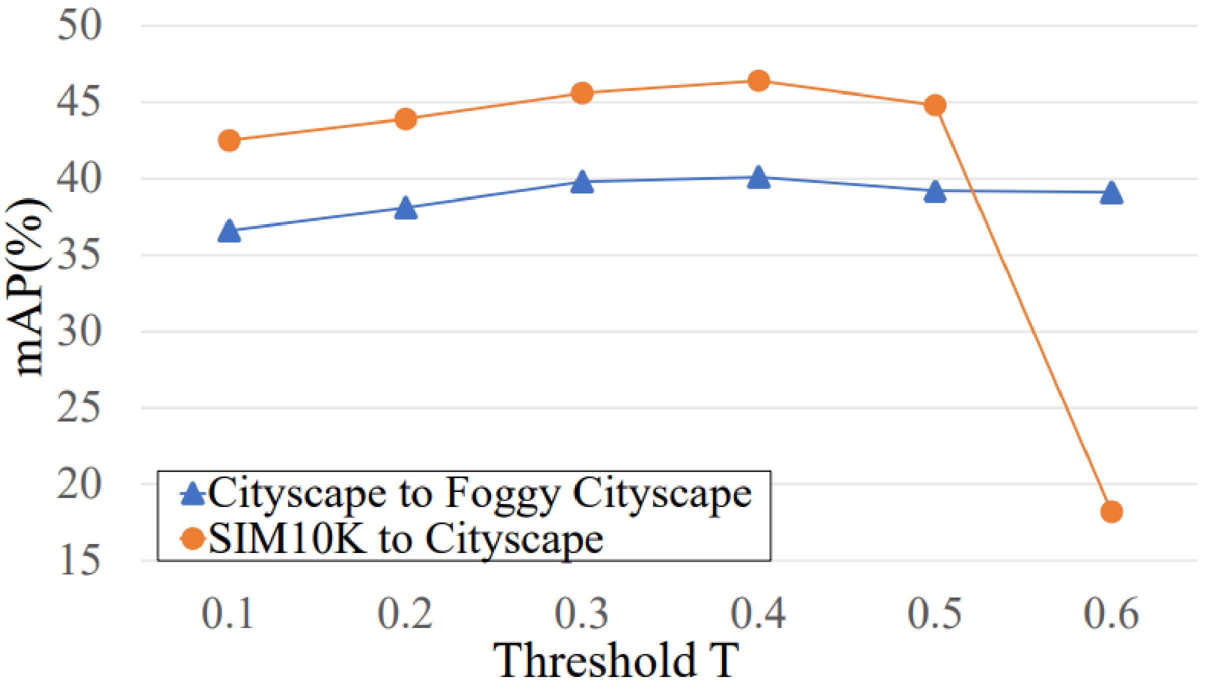
Methods	S to C	K to C
Source trained	34.9	36.5
SCDA	43.0	**42.5**
iFAN	**46.9**	-
MLDA (Ours impl.)	41.8	37.9
Ours (block3,4,5+ins)	46.4	42.0
Oracle	59.1	

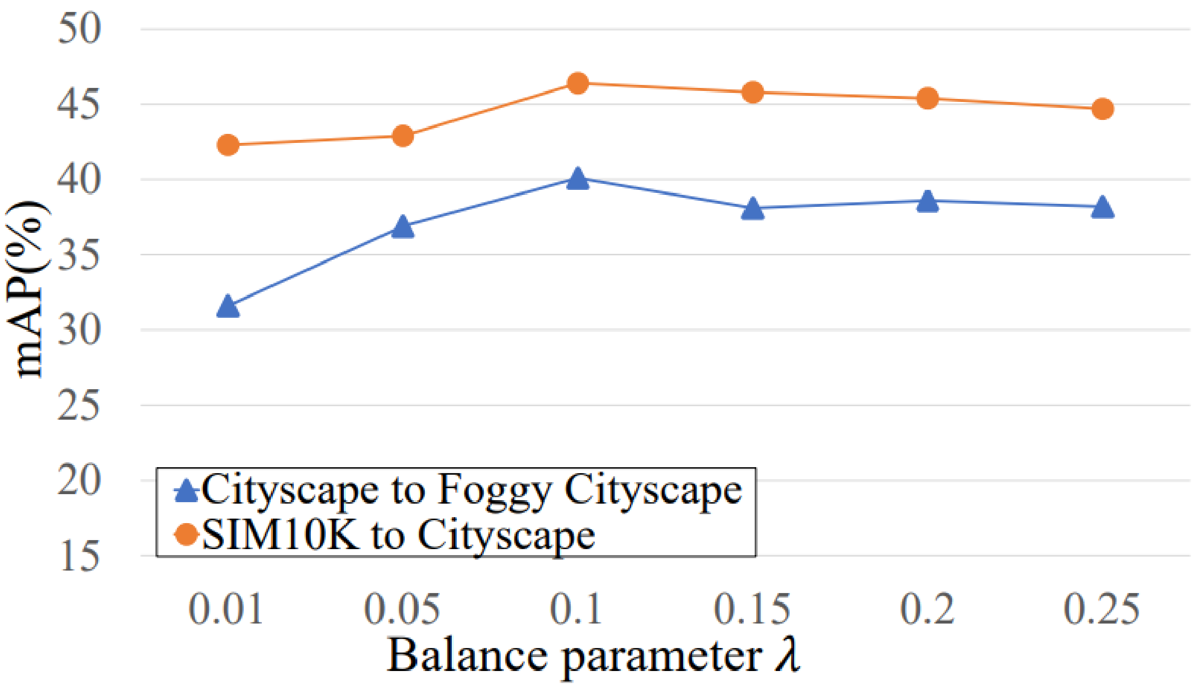
Table 2. Adaptation from BDD100K daytime to nighttime

Methods	bike	bus	car	motor	person	rider	light	sign	train	truck	mAP
Source trained	20.2	33.6	45.7	12.1	27.6	14.0	16.1	31.0	0	30.3	23.1
Strong-Weak	19.6	33.0	46.5	**19.9**	26.4	18.6	15.6	31.5	0	30.9	24.2
MLDA(Our impl.)	20.2	31.8	45.9	16.6	**27.7**	18.2	**16.9**	33.9	0	32.3	24.4
Ours(block3,4,5+ins)	**22.3**	**34.0**	**47.4**	19.7	27.4	**23.0**	14.6	**34.7**	0	**32.7**	**25.6**
Oracle	19.4	39.6	56.1	17.8	29.5	10.9	23	38.9	0	39.1	27.4

Experiments

- Our method has two hyper-parameters: 1) Threshold T for filtering the prediction on target images to provide reliable foreground areas. 2) Parameter λ, which is utilized to balance between the detector loss and adversarial loss.





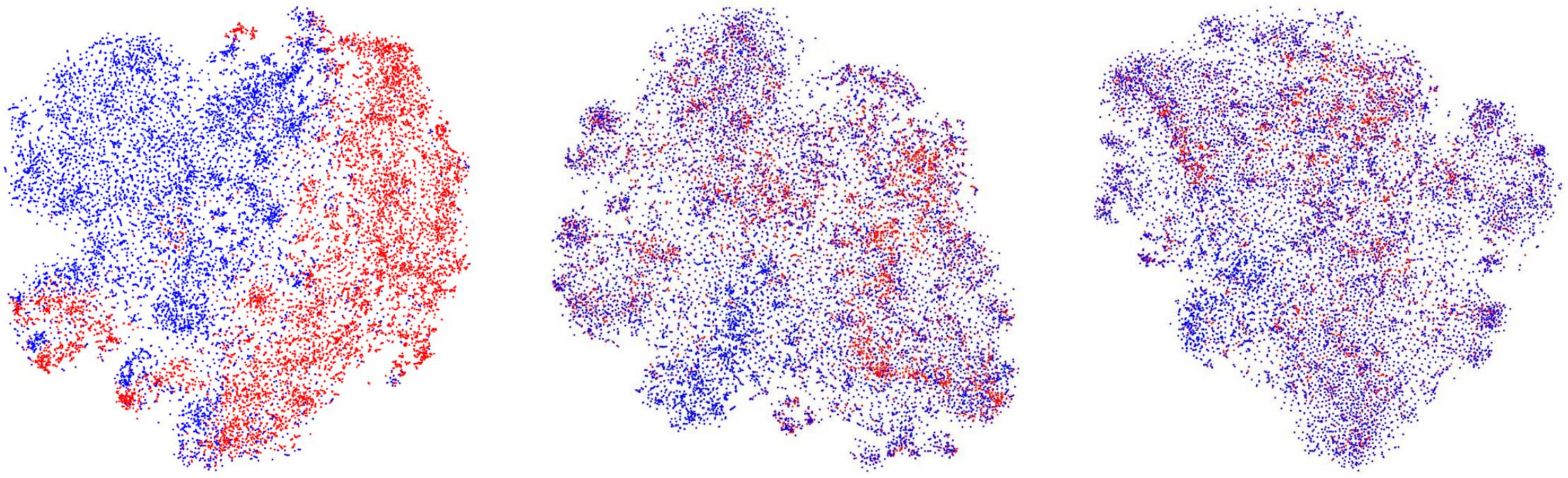
Experiments

- To test the influence of bringing in background adaptation, we replace the FFDA inside our framework with the domain adaptation parts that operate on full feature on different levels as in MLDA.

Methods	C to F	S to C
MLDA (Our impl.)	35.8	41.8
Ours w/ full feature DA on instance level	37.5	45.5
Ours w/ full feature DA on image level block5	39.0	44.5
Ours w/ full feature DA on image level block4	37.2	44.2
Ours w/ full feature DA on image level block3	38.7	44.8
Ours	40.1	46.4

Experiments

- We apply t-SNE on instance level features to observe the feature alignment results visually. Target domain features in blue, source domain in red.



Unadapted MLDA Ours

Conclusion

- We present a straightforward and effective adversarial-based approach for UDA object detection.

- We exploit the crucial factor- ‘foreground adaptation’ that could have significant influence on the adaptation result of object detection.

Adaptation results

1. Clear to foggy weather adaptation
(Cityscape to Foggy Cityscape)

Source domain

Adapt to

Target domain

Clear to foggy weather adaptation

Before adaptation

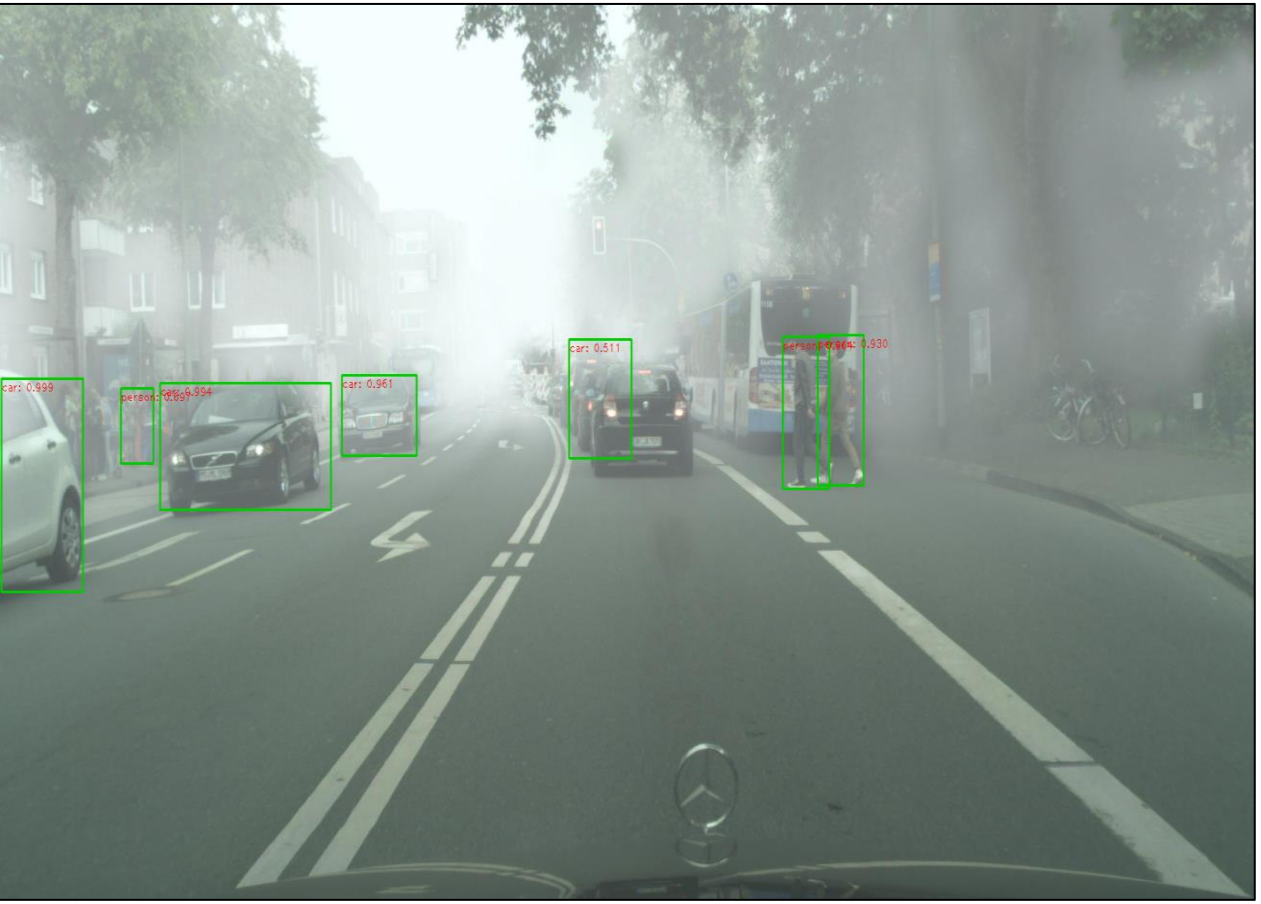
After adaptation

Clear to foggy weather adaptation

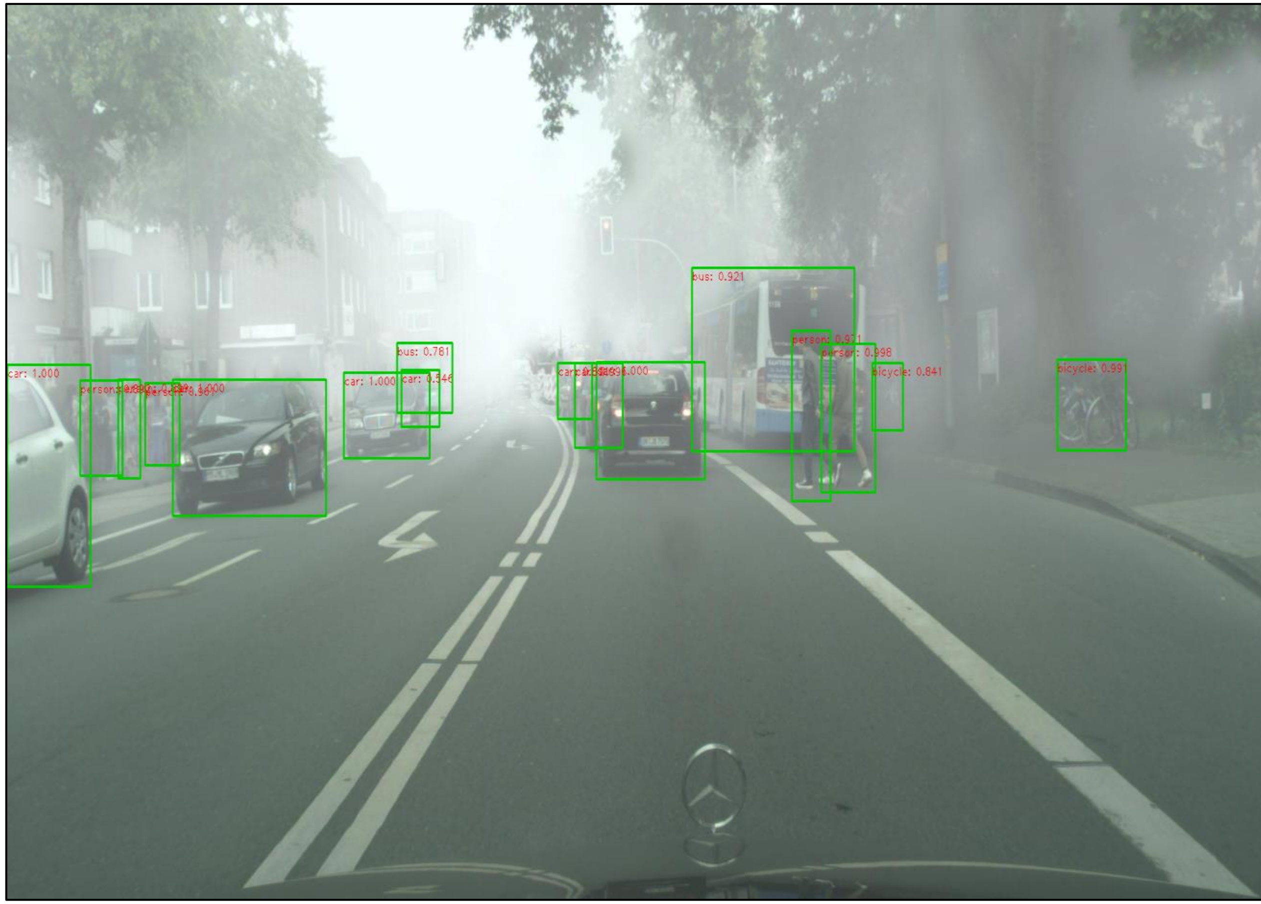
Before adaptation

After adaptation

Clear to foggy weather adaptation



Before adaptation



After adaptation

Adaptation results

2. Synthetic to real adaptation

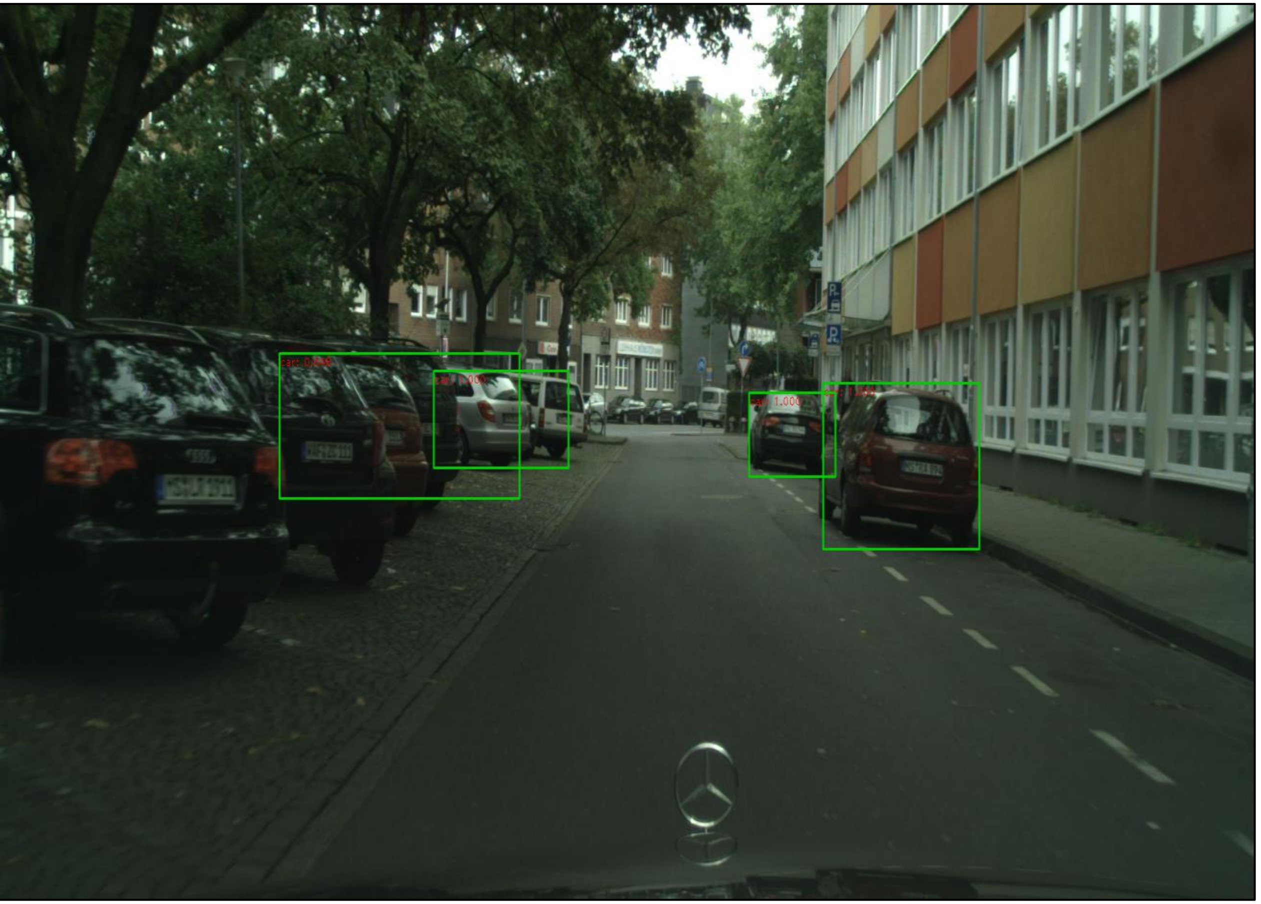
(SIM10K to Cityscape)

Source domain

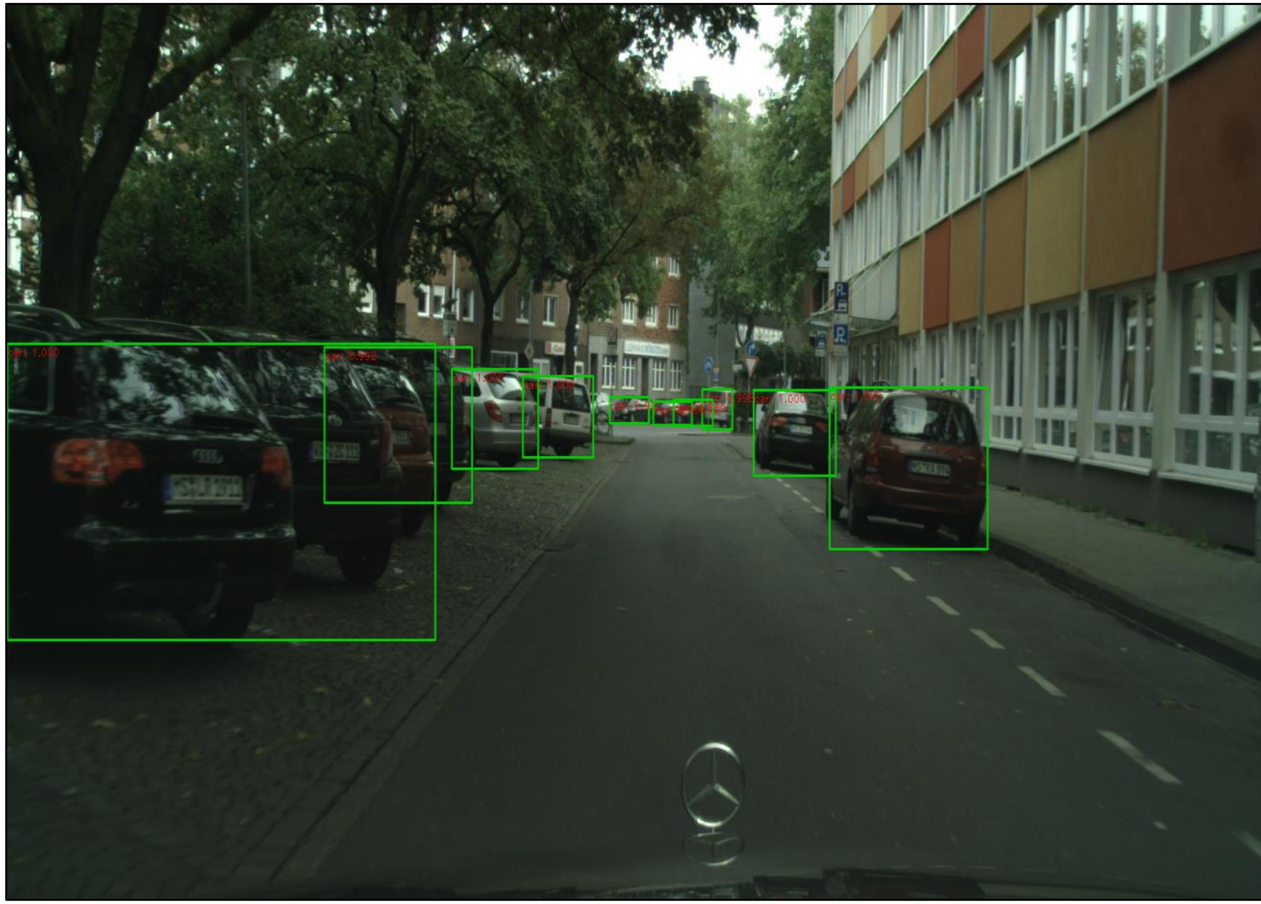
Adapt to

Target domain

Synthetic to real adaptation



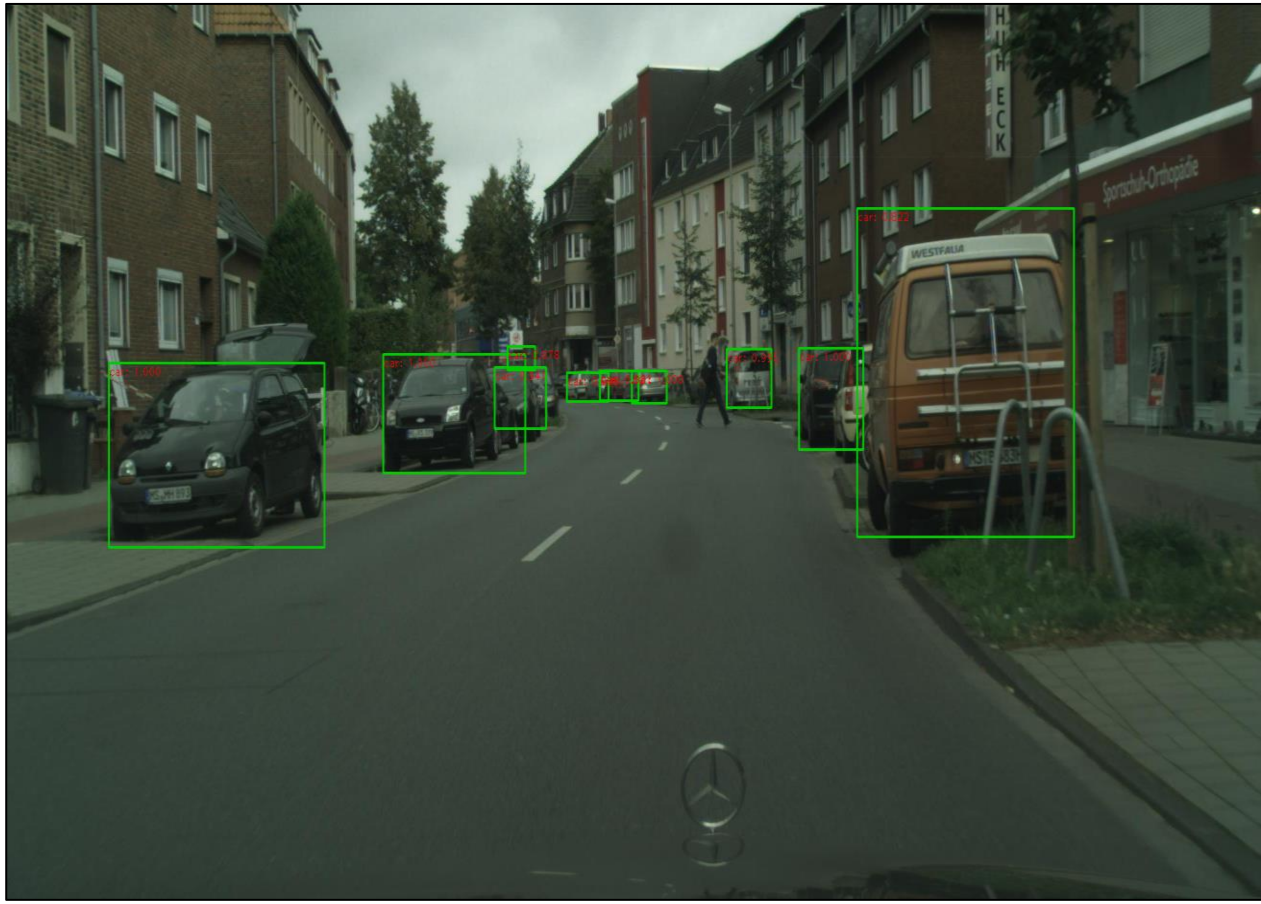
Before adaptation



After adaptation

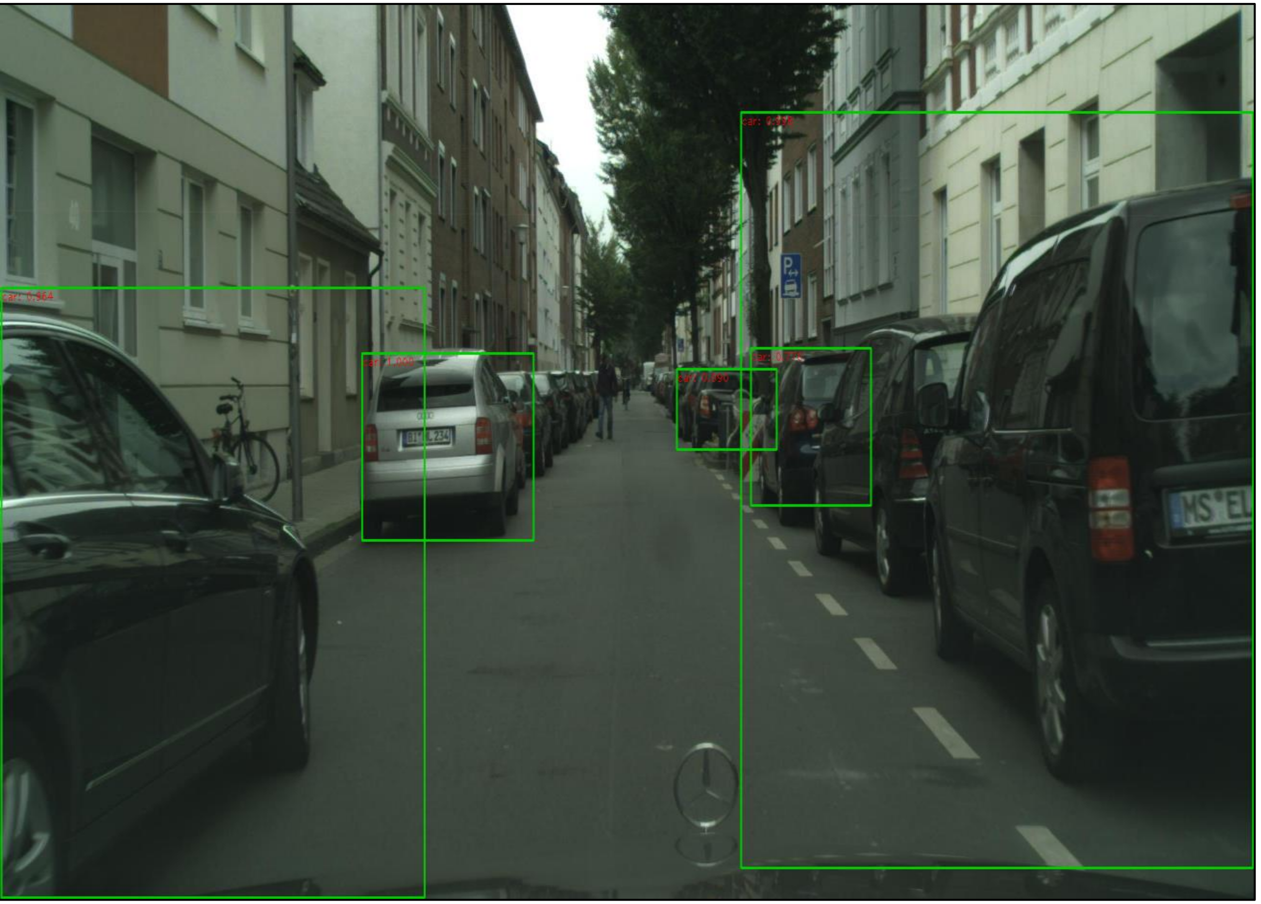
Synthetic to real adaptation

Before adaptation

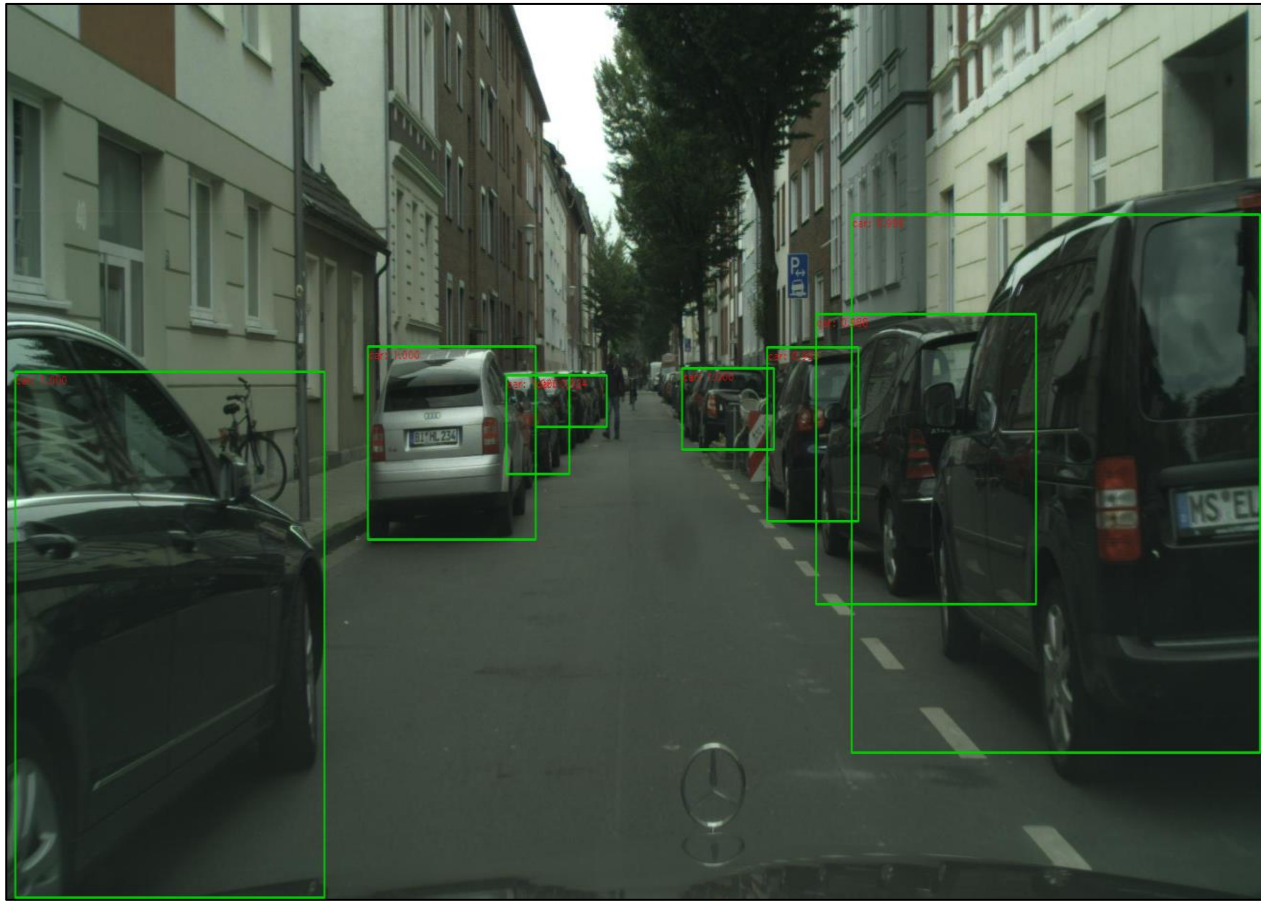


After adaptation

Synthetic to real adaptation



Before adaptation



After adaptation

Adaptation results

3. Cross-camera adaptation (KITTI to Cityscape)

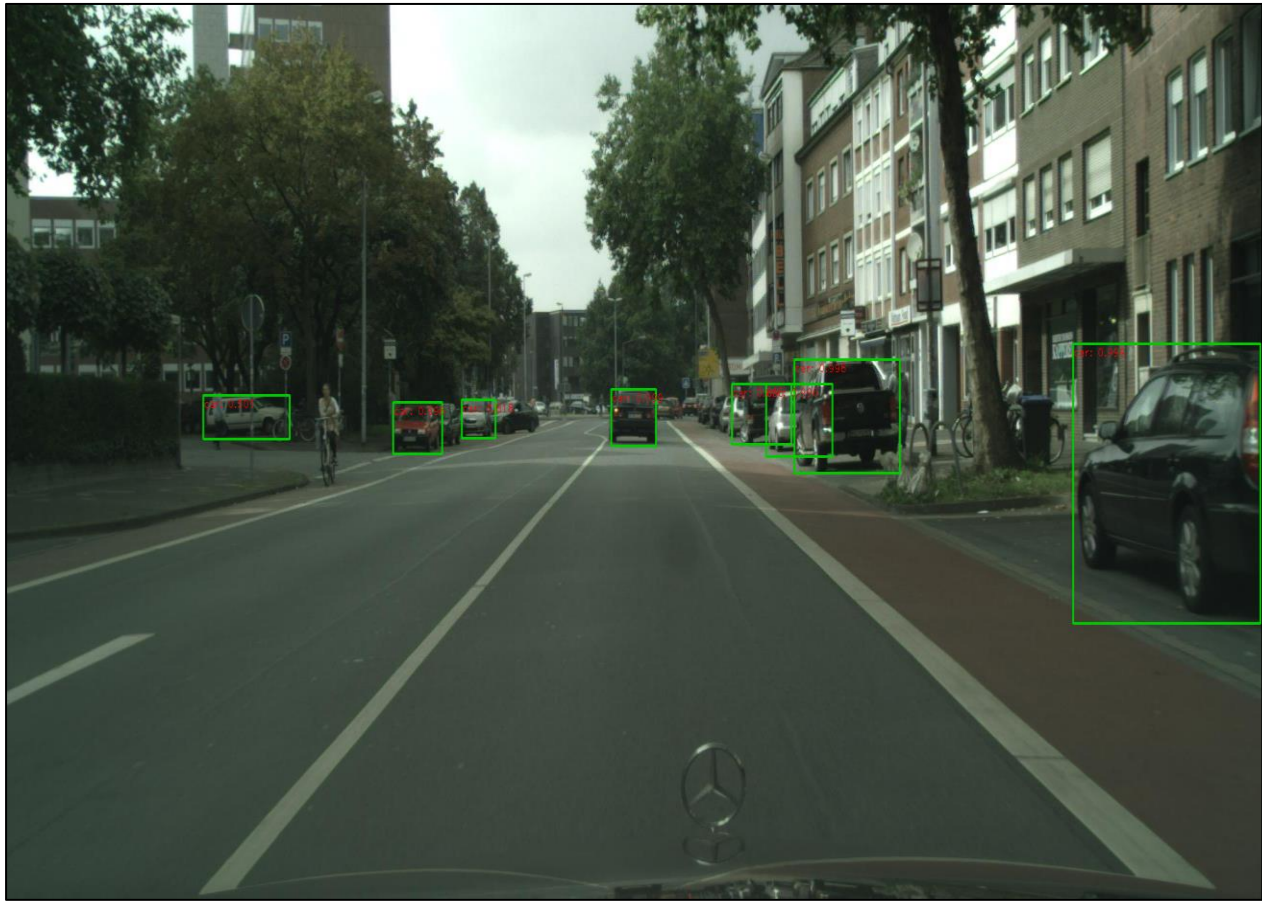
Source domain

Adapt to

Target domain

Cross-camera adaptation

Before adaptation



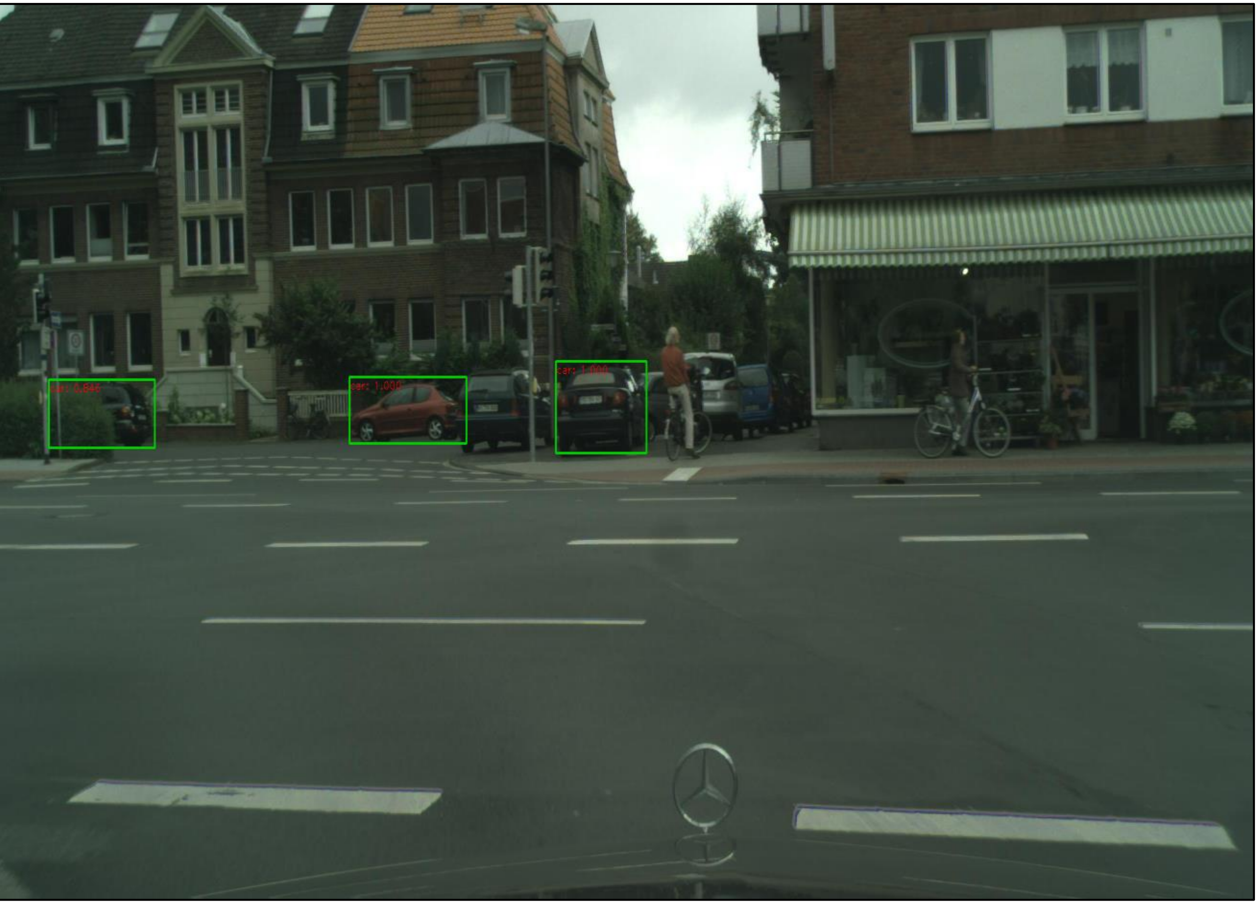
After adaptation

Cross-camera adaptation

Before adaptation

After adaptation

Cross-camera adaptation



Before adaptation

After adaptation

UNIVERSITY OF ALBERTA