Unsupervised Domain adaptation in object detection

Presenter: Yuchen Yang

Supervisor: Nilanjan Ray

Pipeline

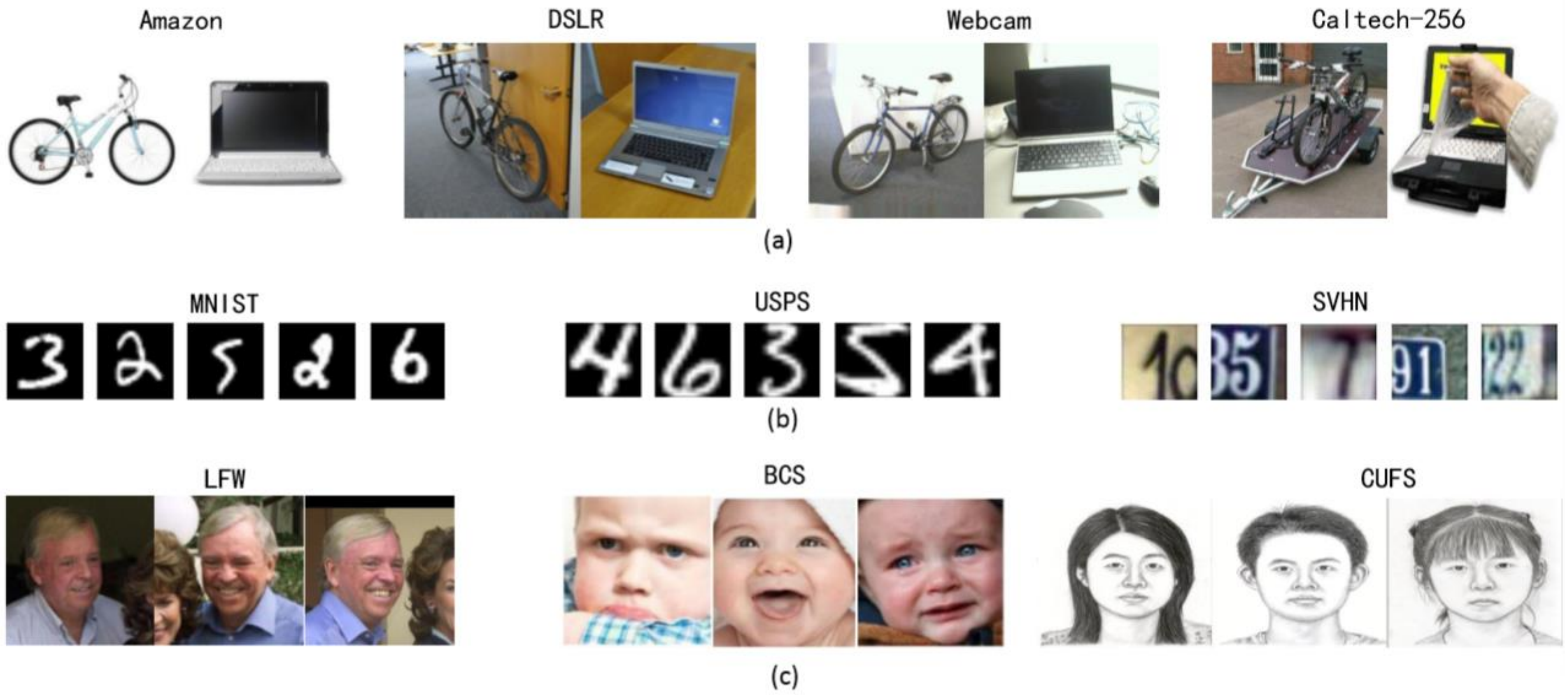
1. Domain adaptation (DA) challenge

2. DA object detection challenge

DA challenge

CNN do not cope well with different distribution shift.

The distribution shift between two datasets degrade the performance.



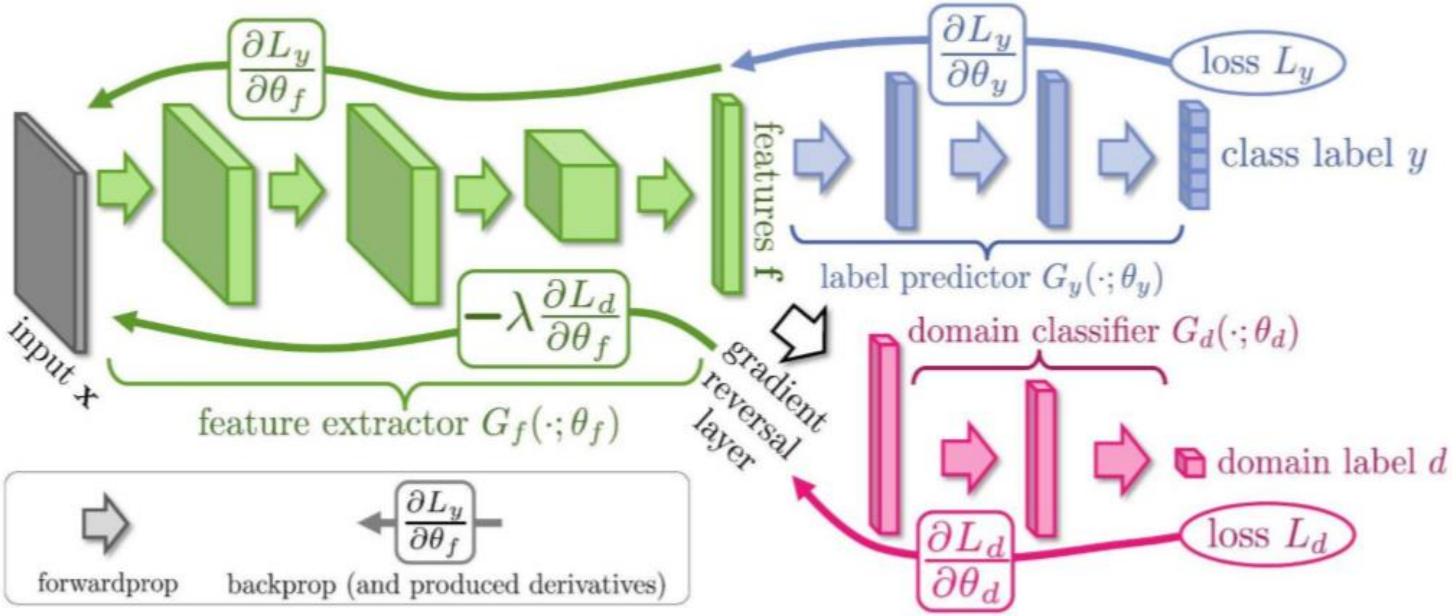
DA challenge

- Domain adaptation is the task that apply a model trained on source domain and apply to target domain.
- Categories:
  - Unsupervised DA
  - Semi-supervised DA
  - Few-shot DA
  - One-shot DA
  - Zero-shot DA
  - ……
- For unsupervised DA, the label in target domain is unavailable. {Xs, Ys} , {Xt}.

DA related work

- Adversarial-based:
  - Domain confusion loss (**DANN**, ADDA)

http://sites.skoltech.ru/compvision/projects/grl/



$$\min_{M^s,C} \mathcal{L}_{cls}(X^s,Y^s) = -\mathbb{E}_{(x^s,y^s)\sim(X^s,Y^s)} \sum_{k=1}^{K} \mathbb{1}_{[k=y^s]} \log C(M^s(x^s))$$

$$\min_{D} \mathcal{L}_{advD}(X^s,X^t,M^s,M^t) = -\mathbb{E}_{(x^s)\sim(X^s)}[\log D(M^s(x^s))] - \mathbb{E}_{(x^t)\sim(X^t)}[\log(1-D(M^t(x^t)))]$$

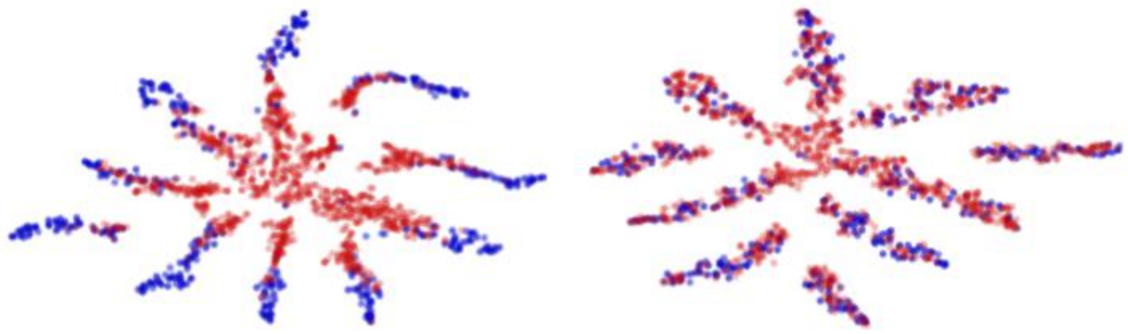
Unsupervised Domain Adaptation by Backpropagation (DANN). ICML2015

T-SNE plots before and after DA

MNIST → MNIST-M: top feature extractor layer

(a) Non-adapted (b) Adapted

SYN NUMBERS → SVHN: last hidden layer of the label predictor



(a) Non-adapted (b) Adapted

DA object detection challenge

Cityscape

Foggy cityscape
(different weather)

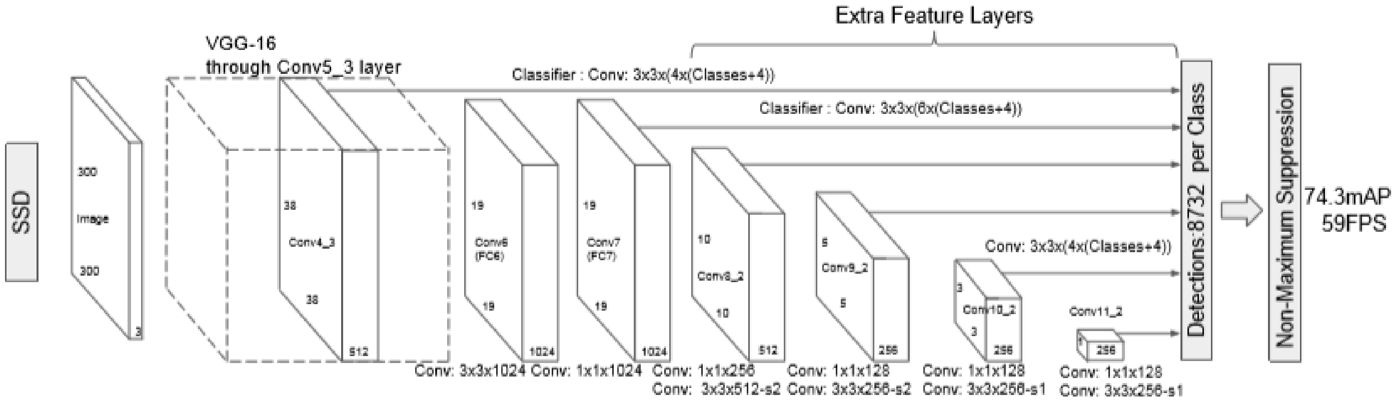
Sim10k
(simulated data)

Kitti
(Camera & scale
variances)

Object detection background

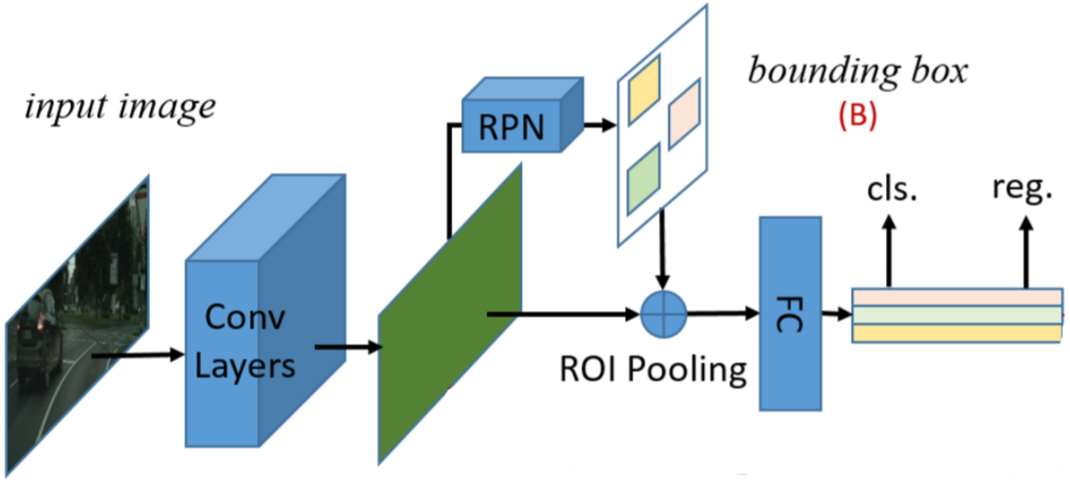
One-stage object detectors

SSD, YOLO, Retinanet



Two-stage object detectors

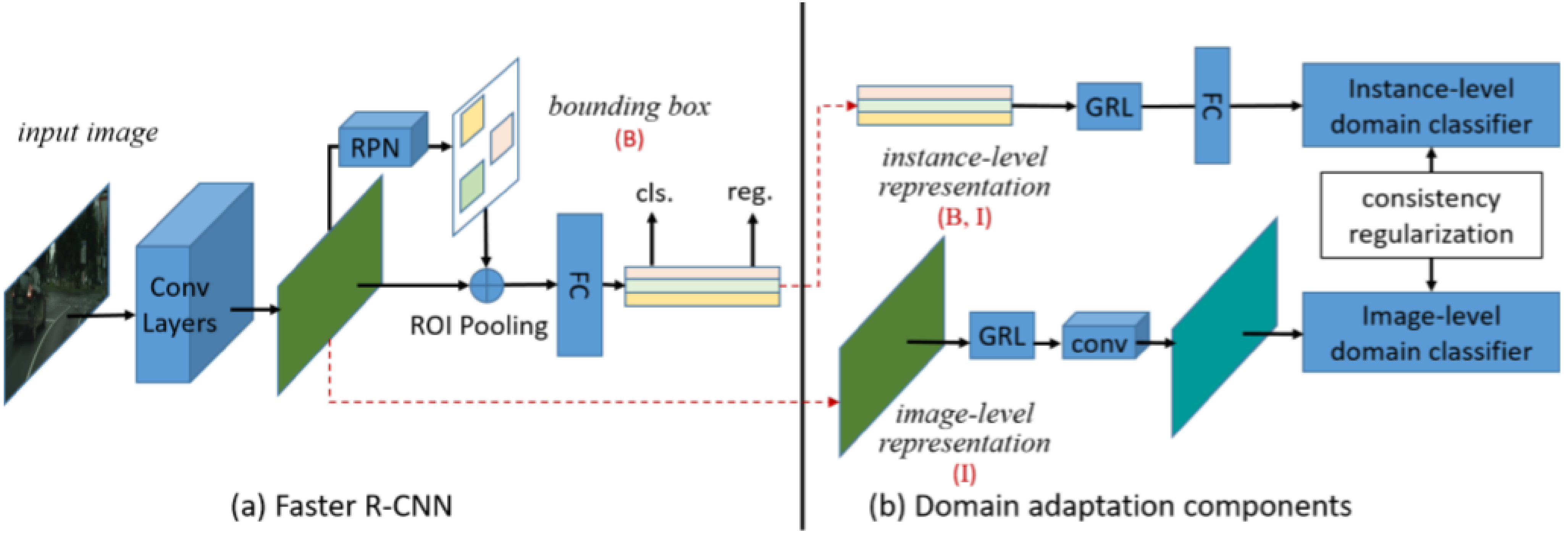
Faster rcnn, rfcn, mask-rcnn



DA object detection related work

- Adversarial Feature representation methods:
  - Domain Adaptive Faster R-CNN for Object Detection in the Wild. CVPR2018
  - Strong-Weak Distribution Alignment for Adaptive Object Detection. CVPR2019
  - Adapting Object Detectors via Selective Cross-Domain Alignment. CVPR2019
  - Multi-adversarial Faster-RCNN for Unrestricted Object Detection. ICCV2019
  - Multi-Level Domain Adaptive Learning for Cross-Domain Detection. ICCV workshop 2019
  - SCL: Towards Accurate Domain Adaptive Object Detection via Gradient Detach Based Stacked Complementary Losses. Arxiv Nov 21

DA faster rcnn



$$\mathcal{L}_{img} = -\sum_{i,u,v}\left[D_i \log p_i^{(u,v)} + (1-D_i)\log(1-p_i^{(u,v)})\right]$$

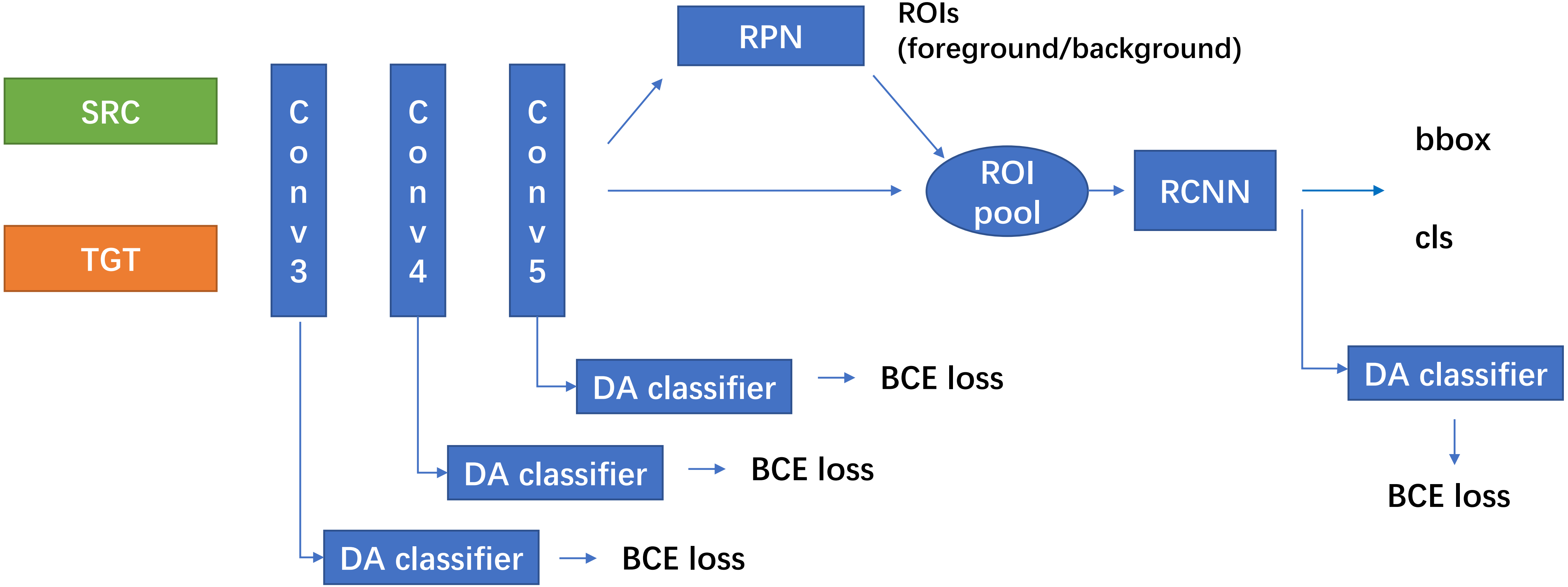
$$\mathcal{L}_{ins} = -\sum_{i,j}\left[D_i \log p_{i,j} + (1-D_i)\log(1-p_{i,j})\right]$$

$$L_{cst} = \sum_{i,j}\|\frac{1}{|I|}\sum_{u,v}p_i^{(u,v)} - p_{i,j}\|_2$$

$$L = L_{det} + \lambda(L_{img} + L_{ins} + L_{cst})$$

Domain Adaptive Faster R-CNN for Object Detection in the Wild. CVPR 2018

Multi-layer DA faster rcnn



Results

1. Different weather condition. (Cityscape -> foggy cityscape)

froggy cityscape -> cityscape	person	rider	car	truck	bus	train	mcycle	bicycle	mAP
Faster RCNN	24.1	25.7	27.2	10.4	17.0	9.1	16.7	25.4	19.4
DA faster rcnn (CVPR 2018)	25.0	31.0	40.5	22.1	35.3	20.2	20.0	27.1	27.6
MAF (ICCV 2019)	28.2	39.5	43.9	23.8	39.9	33.3	29.2	33.9	34
Multi-da faster (ICCV workshop 2019)	33.2	44.2	44.8	28.2	41.8	28.7	30.5	36.5	36.0
Strong-Weak CVPR2019	30	40	43.4	23.2	40.1	34.6	27.8	33.4	34.1
CMU-SCL (arxiv Nov 21)	31.6	44	44.8	30.4	41.8	40.7	33.6	36.2	37.9
Microsoft Nov 24 (DT only)	36.5	41.7	54.6	22.6	40.7	25.3	29.3	36.9	35.9
Microsoft Nov 24 (DT + ST)	38.2	42.1	55.6	25.9	43.5	27.6	33.5	39.2	38.2
Upper Bound	33.2	45.9	49.7	35.6	50	37.4	34.7	36.2	40.3

Results

2. From simulated data to real data.
(Sim10k to cityscape)

	mAP
Faster RCNN	34.6
DA faster rcnn (CVPR 2018)	38.9
Strong-weak (CVPR 2019)	40.1
Multi-da faster (ICCV workshop 2019)	42.0
CMU-SCL (Nov 21)	42.6
SC-DA 512 (CVPR 2019)	43.0
Microsoft Nov 24 (DT only)	50.8
Microsoft Nov 24 (DT +ST)	52.3

3. Cityscape to Kitti

(Camera & scale shift)

	mAP
Faster RCNN	30.2
DA faster rcnn (CVPR 2018)	38.5
Strong-weak (CVPR 2019)	37.9
CMU-SCL (Nov 21)	41.9
Microsoft Nov 24 (DT only)	42.8
Microsoft Nov 24 (DT +ST)	46.4

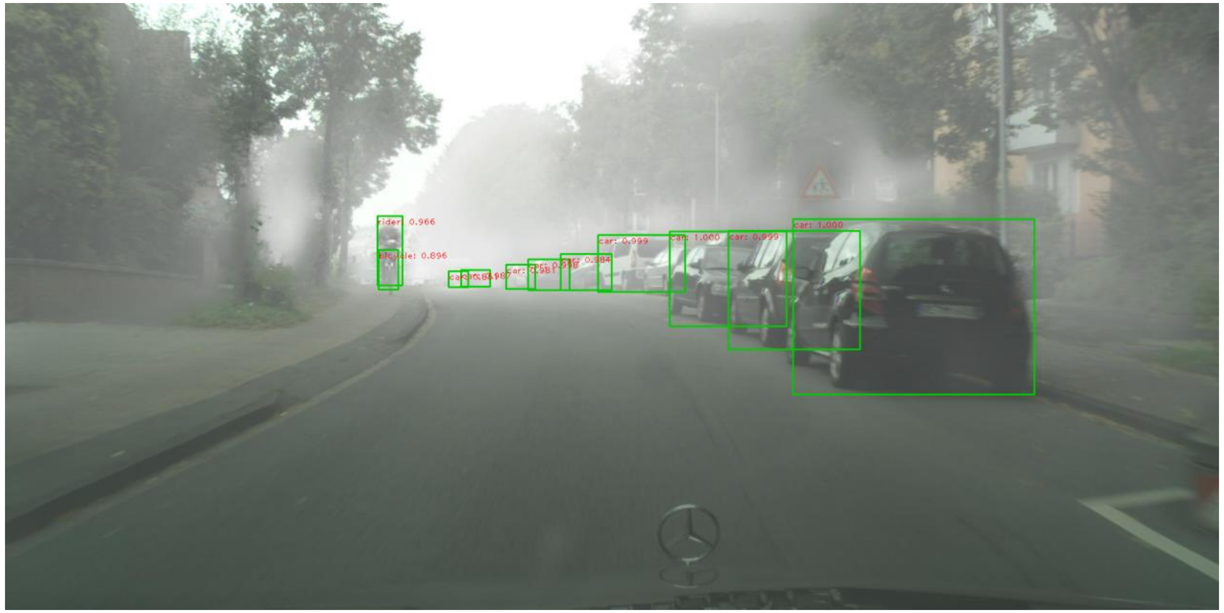
Results (cityscape -> foggy cityscape)

Without adaptation

With adaptation

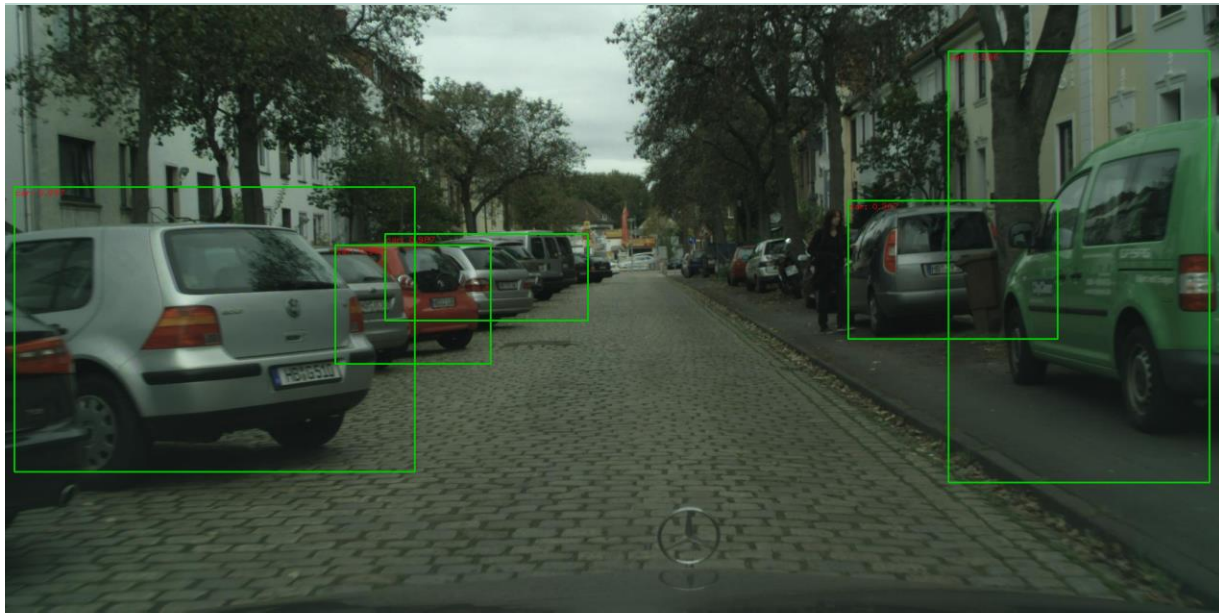
Results (cityscape -> foggy cityscape)

Without adaptation

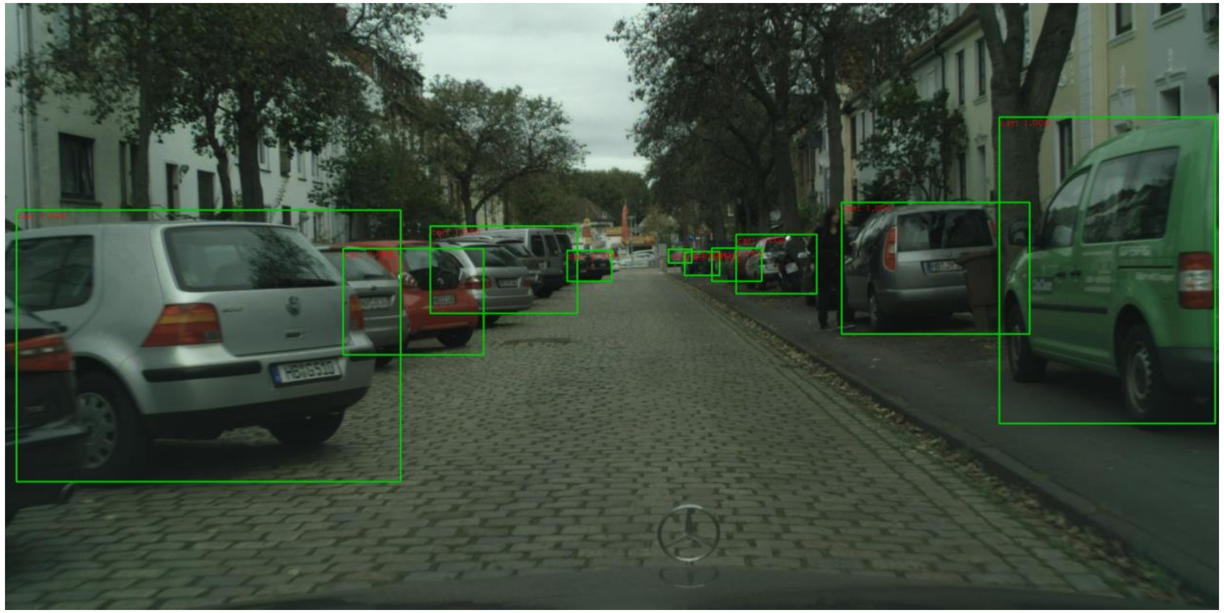


With adaptation

Results (sim10k -> cityscape)

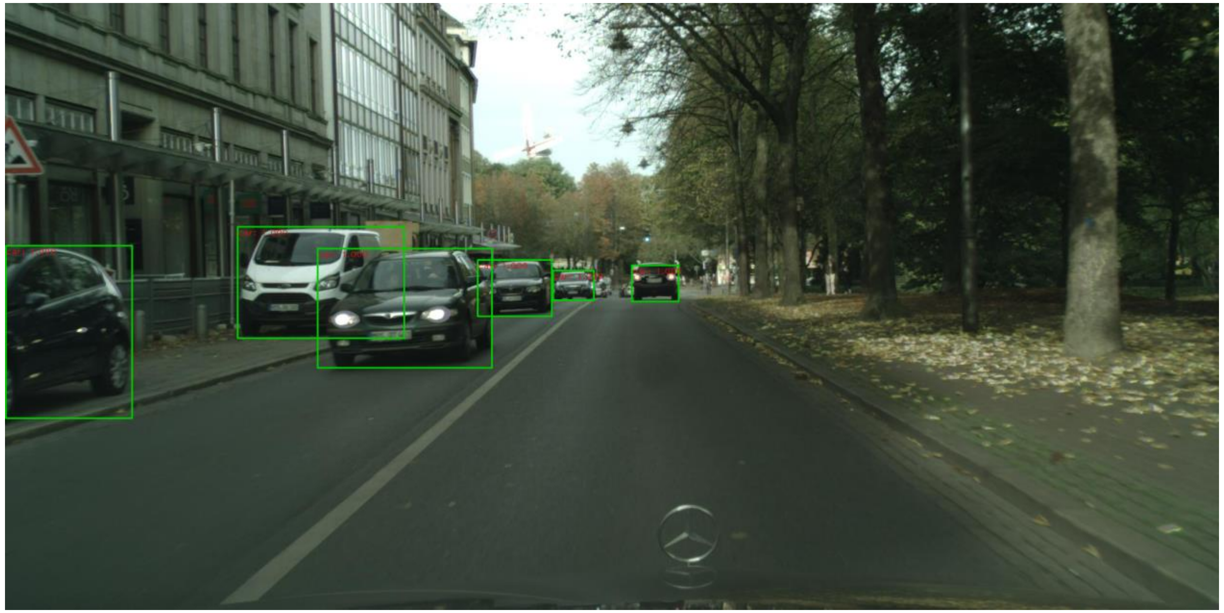


Without adaptation



With adaptation

Results (sim10k -> cityscape)



Without adaptation

With adaptation

Foreground-focused domain adaptation for object detection

YUCHEN WANG, NILANJAN RAY

PAPER AT ICPR 2020

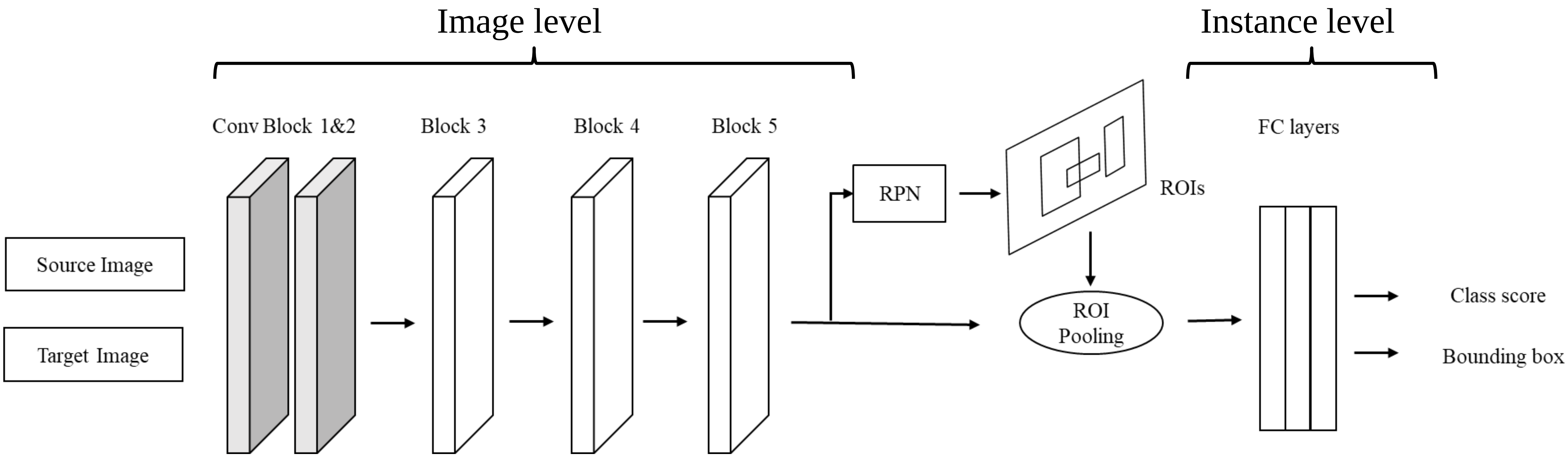
Introduction

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.

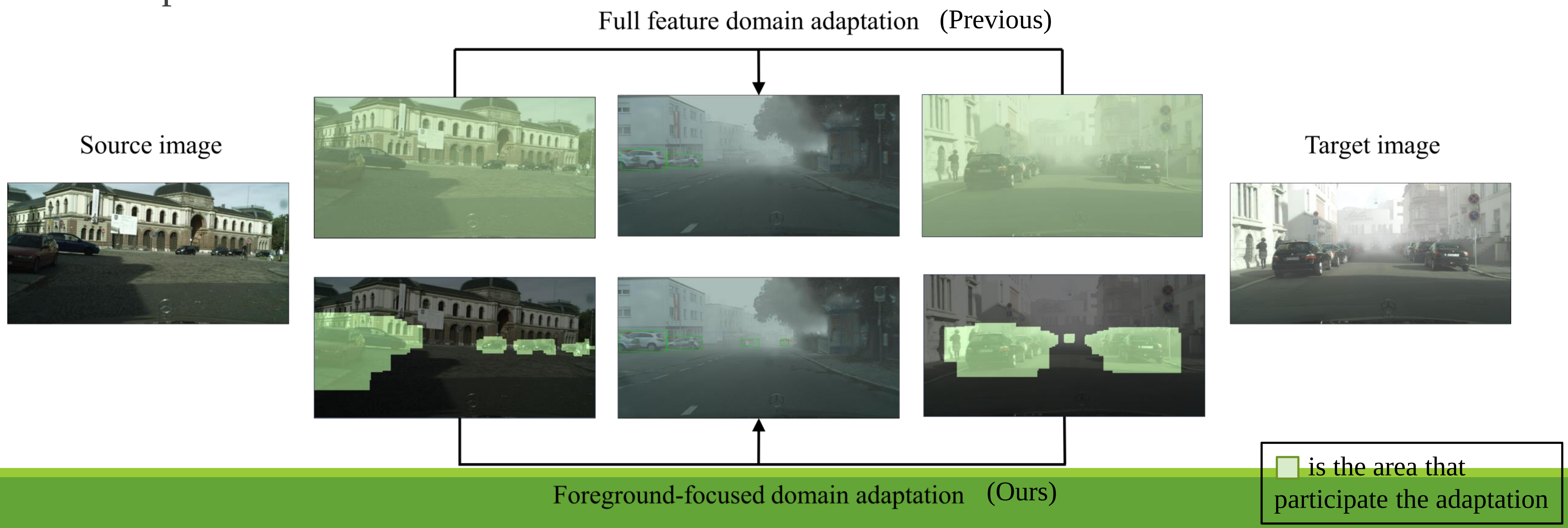
Preliminary

We use Faster RCNN as the detector for adaptation. Faster RCNN can be divided to two parts: Image level and instance level.



Method

We propose Foreground-focused domain adaptation (FFDA). We use mining masks in different feature levels that allow only foreground areas to participate in adaptation.



Mining masks generation

Mining masks are generated using source ground-truth and target predictions.

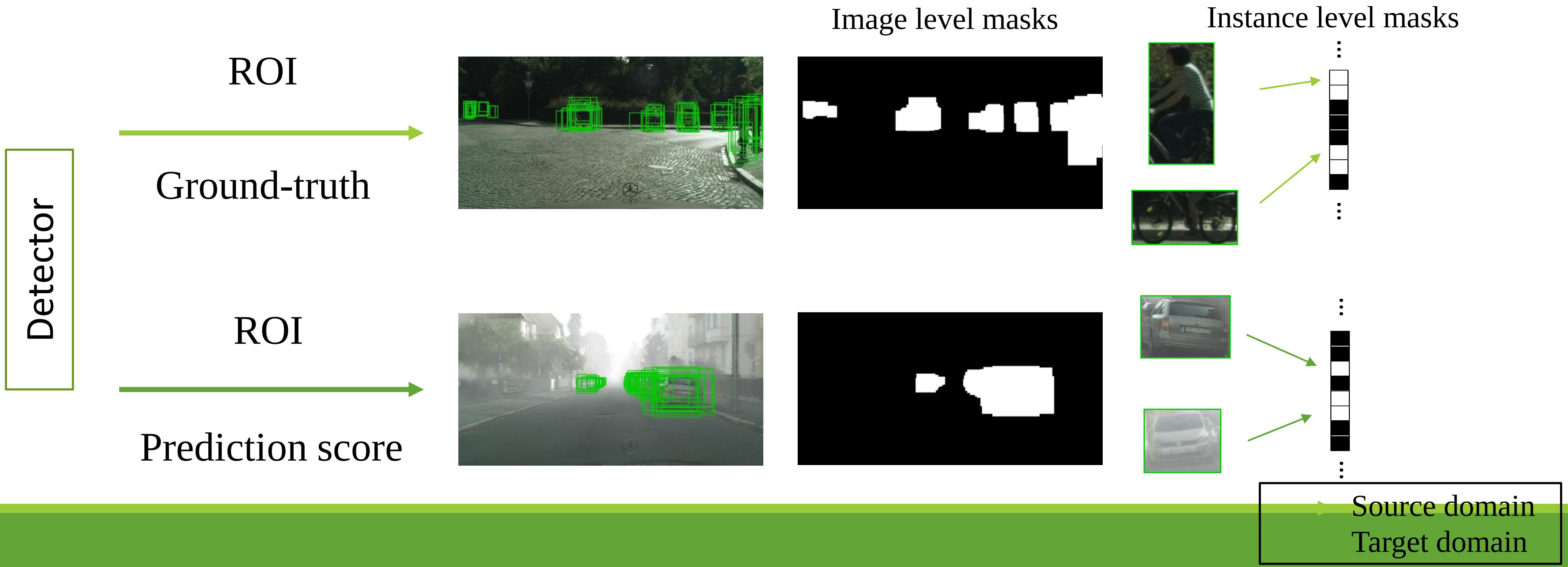
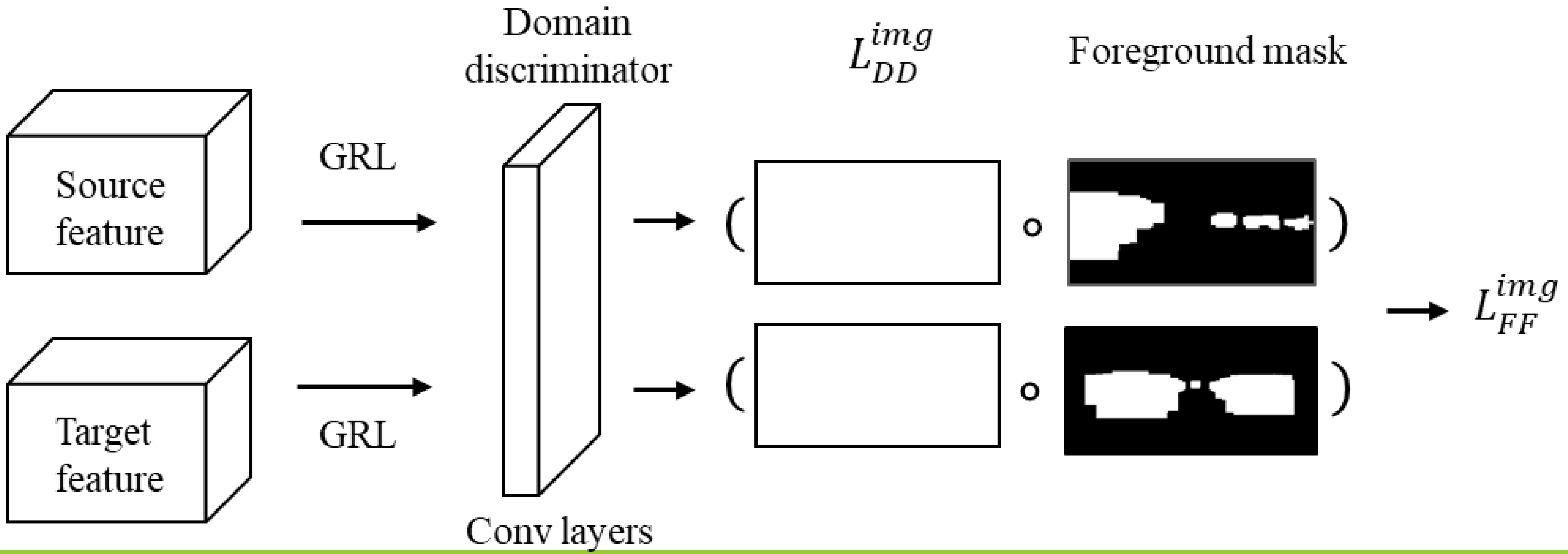


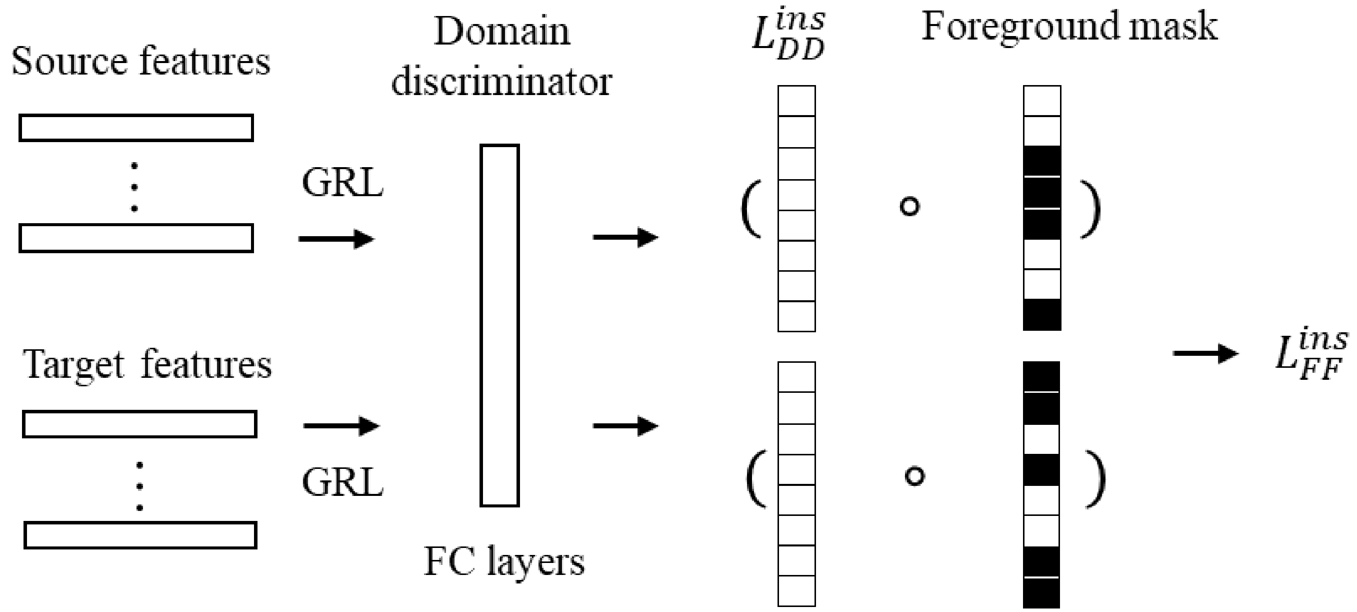
Image level FFDA

We mine the loss in foreground area on the loss map generated from image level domain discriminator.



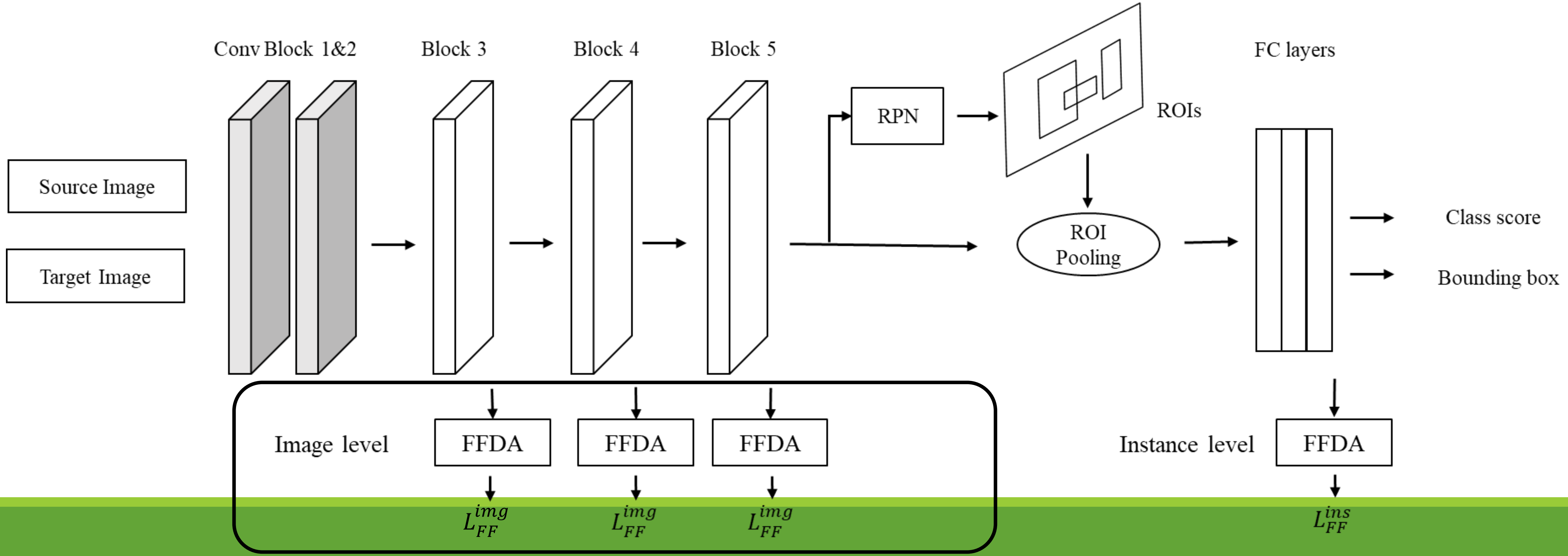
Instance level FFDA

We mine the loss of identified foreground ROI features from instance level domain discriminator.



Multi-adversarial alignment

We apply multi-adversarial alignment on Image level to build strong alignment on features.



Adaptation results

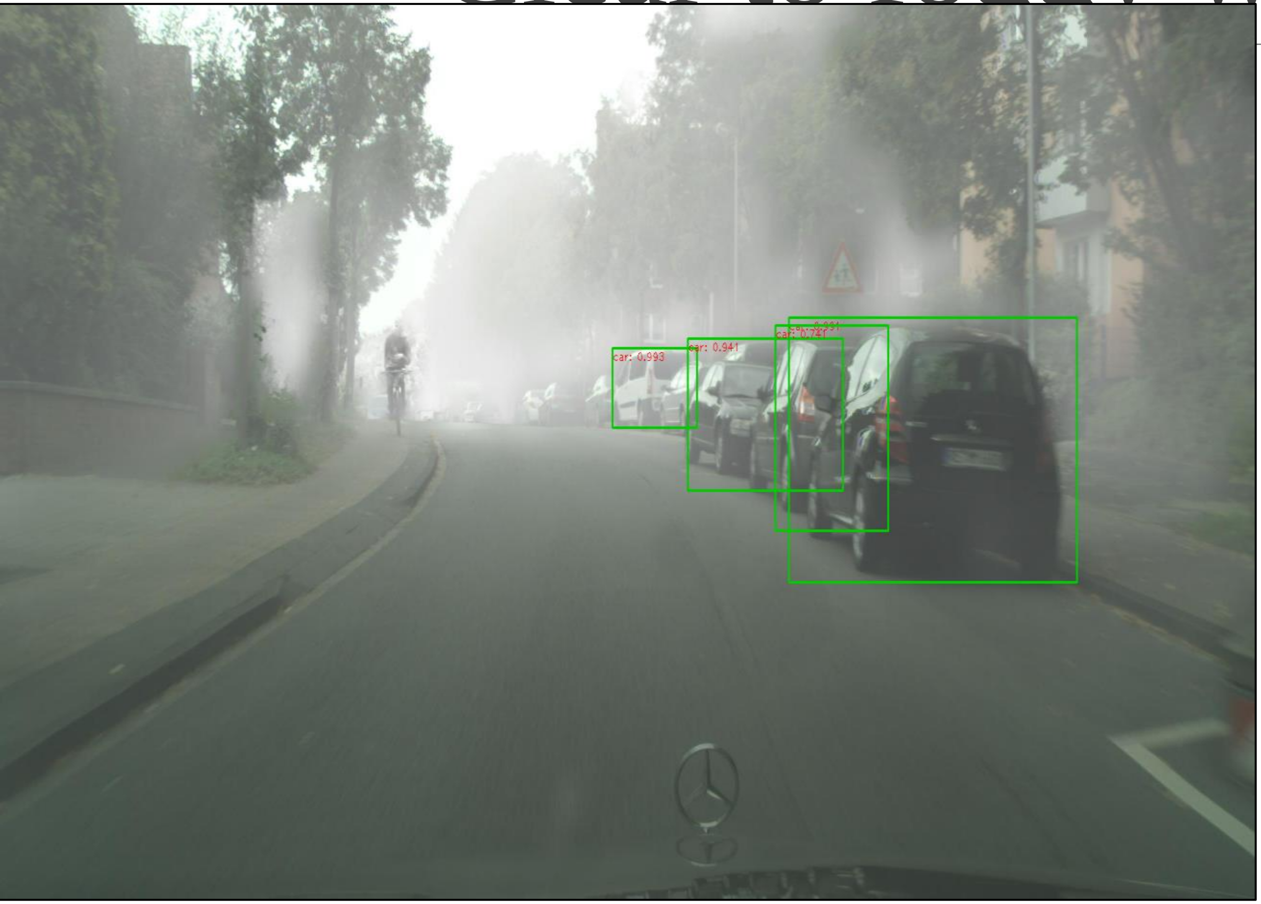
1. CLEAR TO FOGGY WEATHER ADAPTATION (CITYSCAPE TO FOGGY CITYSCAPE)

Source domain

Adapt to

Target domain

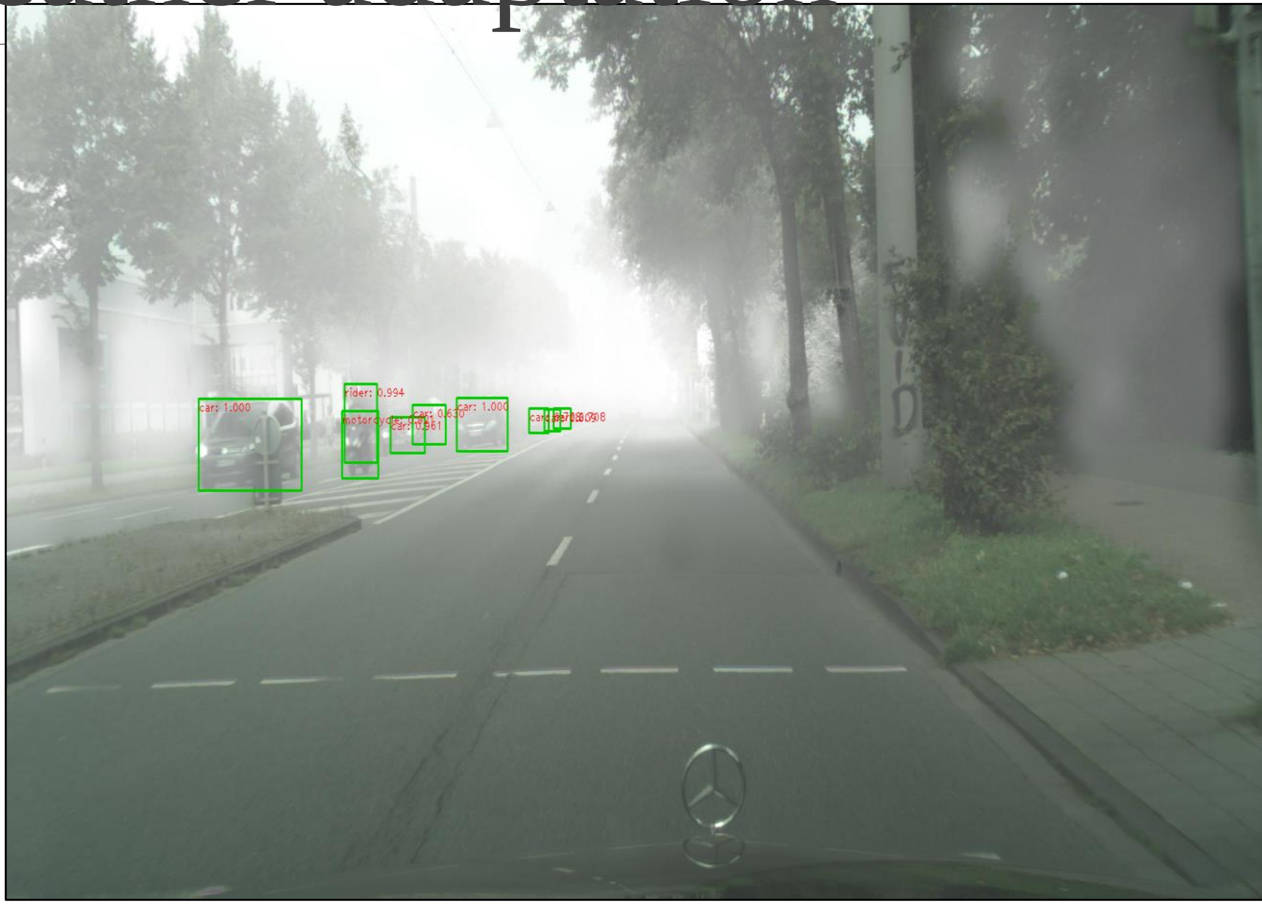
Clear to foggy weather adaptation



Source trained (Non-adapted)

Ours

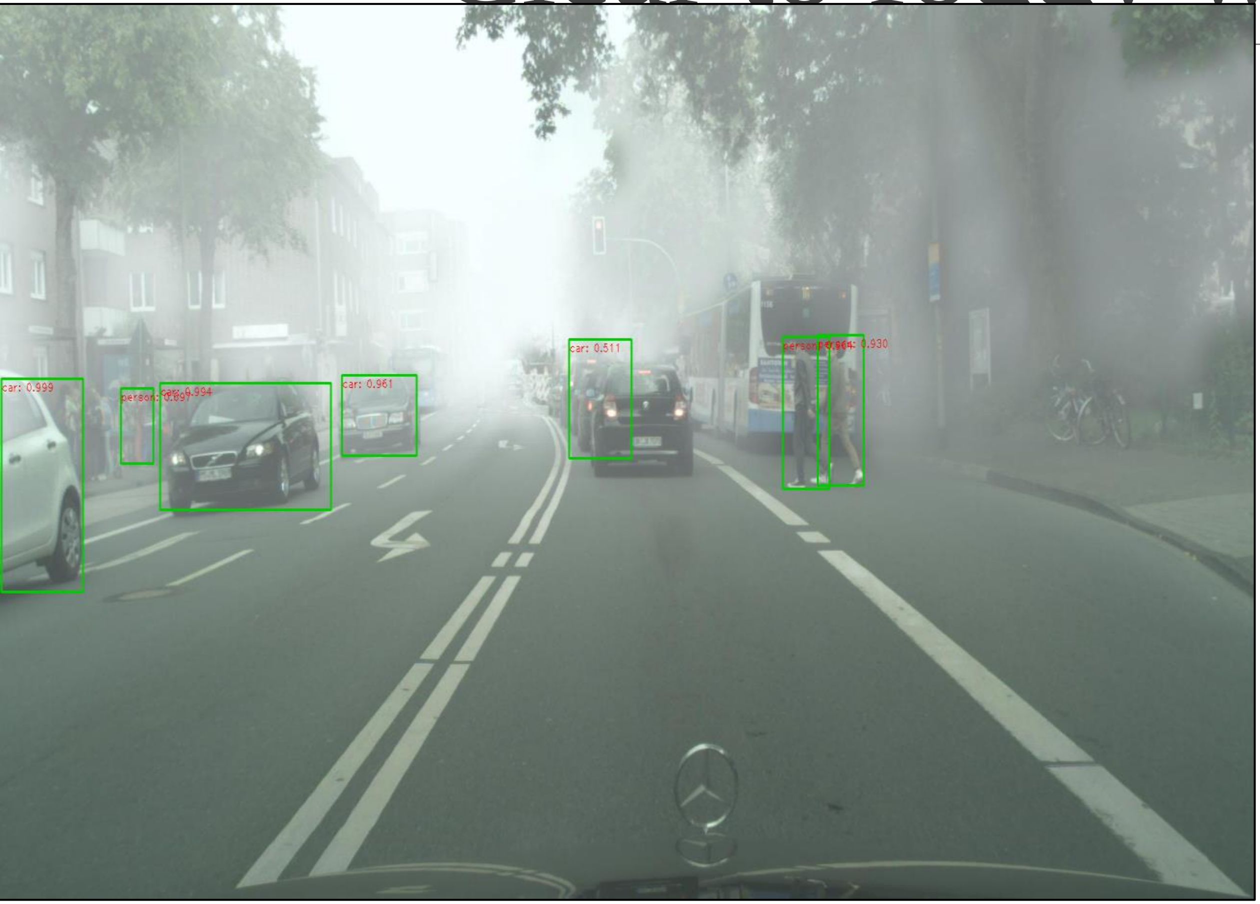
Clear to foggy weather adaptation

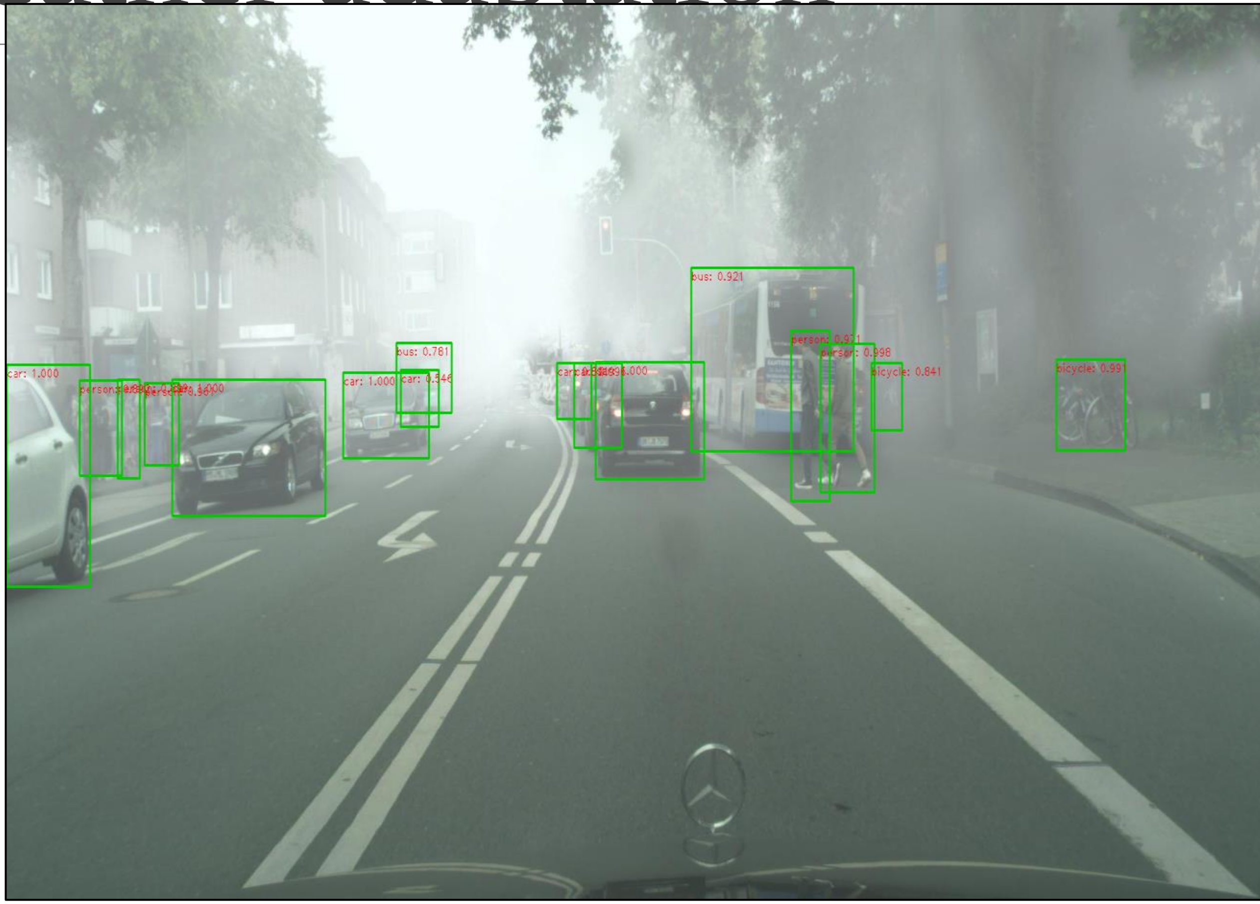


Source trained (Non-adapted)

Ours

Clear to foggy weather adaptation

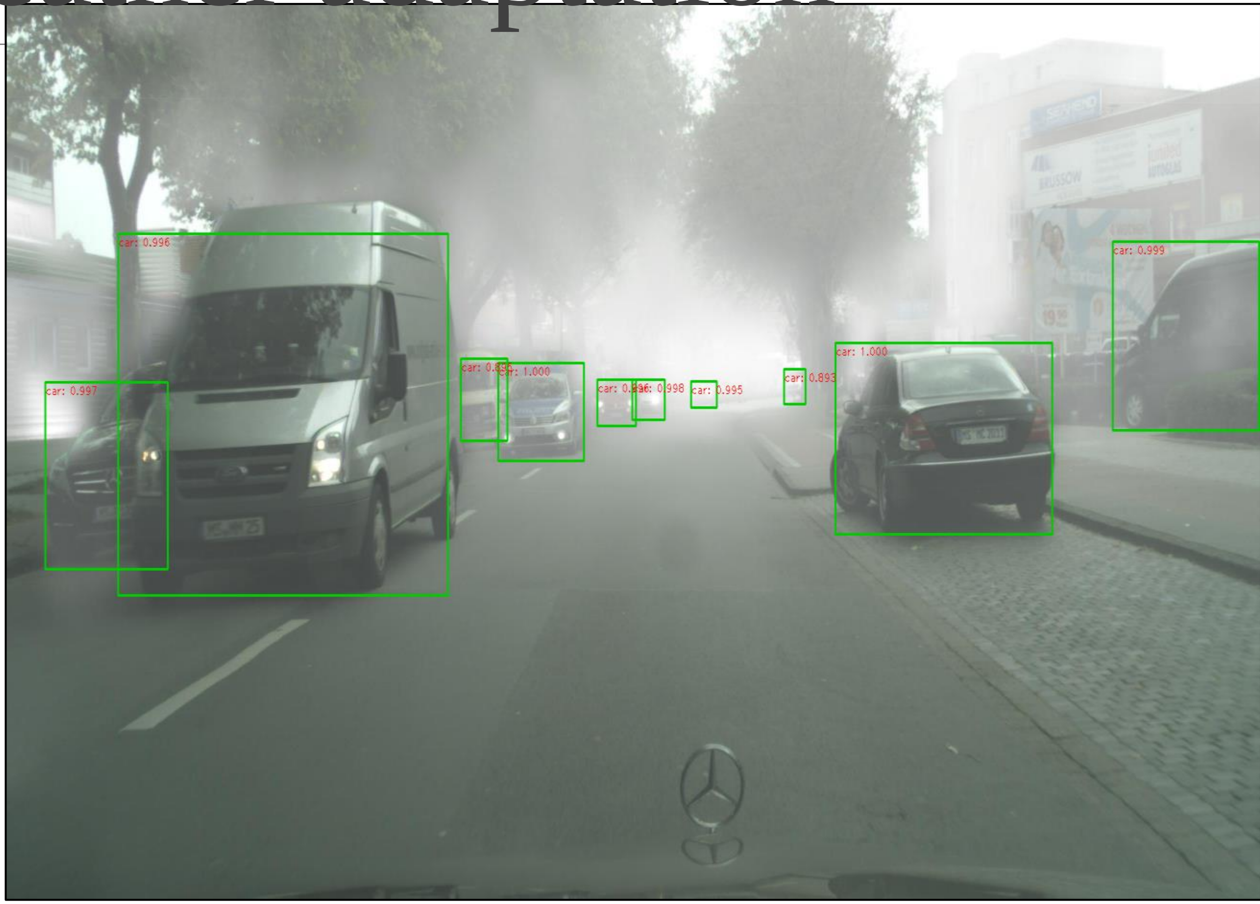




Source trained (Non-adapted) Ours

Clear to foggy weather adaptation

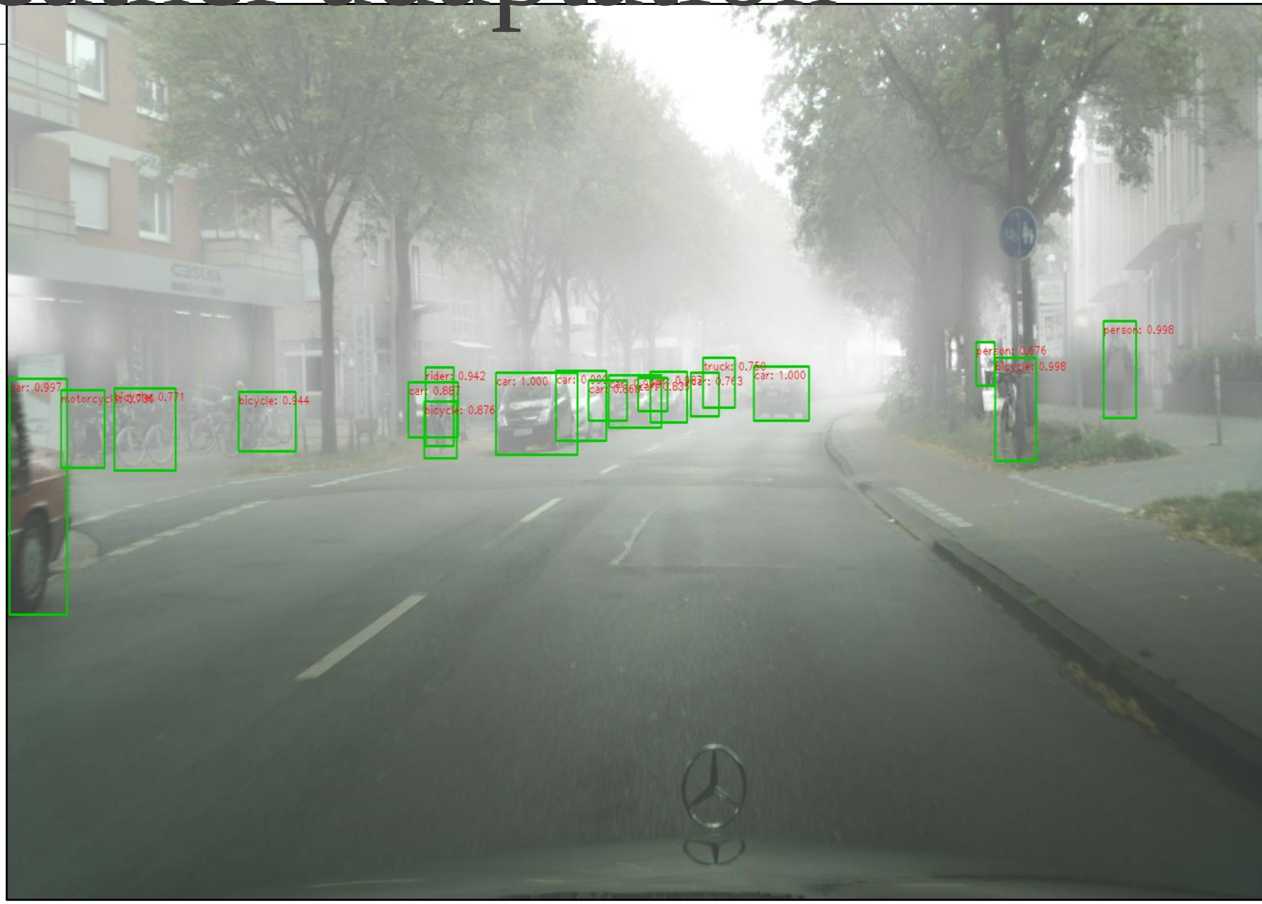
Source trained (Non-adapted)



Ours

Clear to foggy weather adaptation

Source trained (Non-adapted)



Ours

Adaptation results

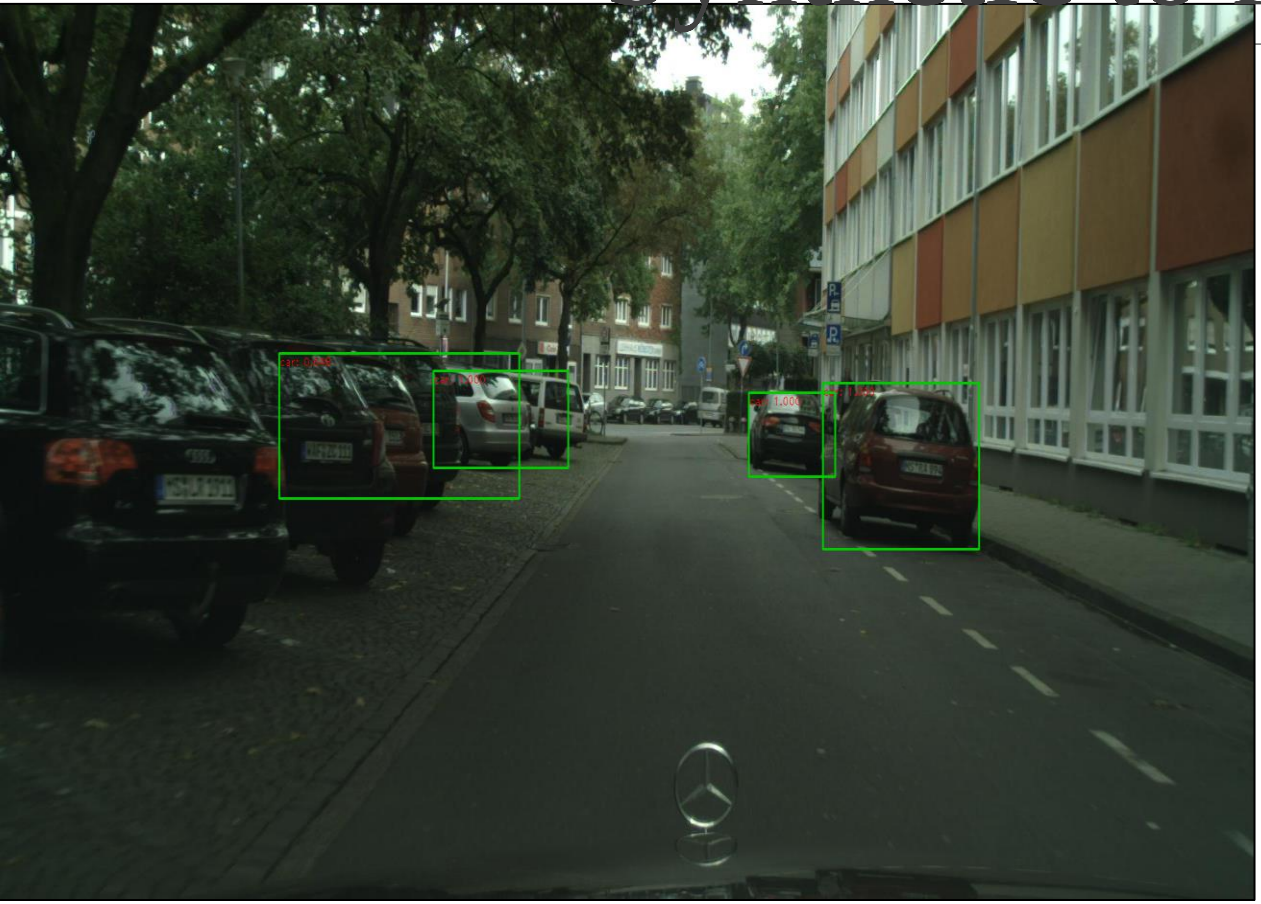
2. SYNTHETIC TO REAL ADAPTATION (SIM10K TO CITYSCAPE)

Source domain

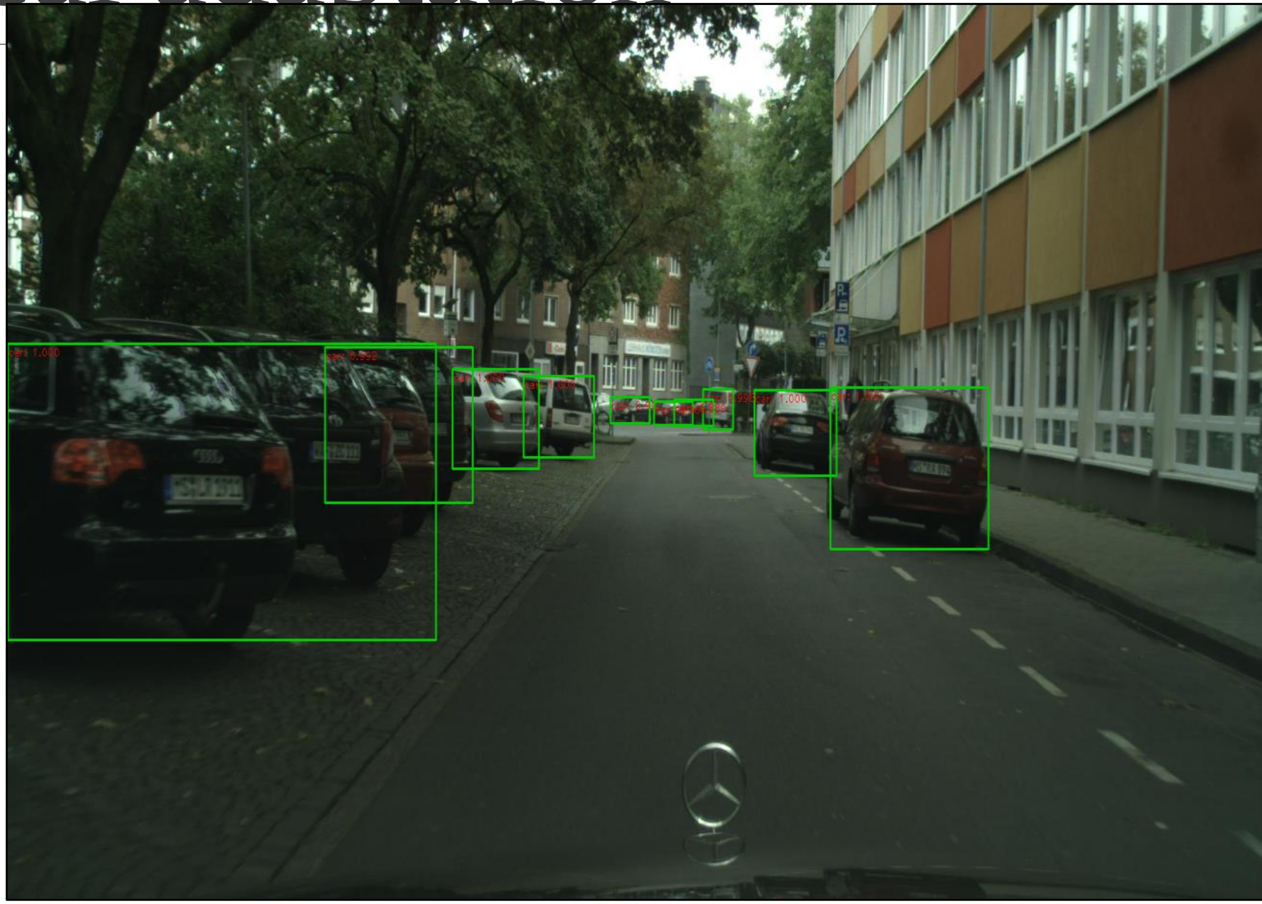
Adapt to

Target domain

Synthetic to real adaptation



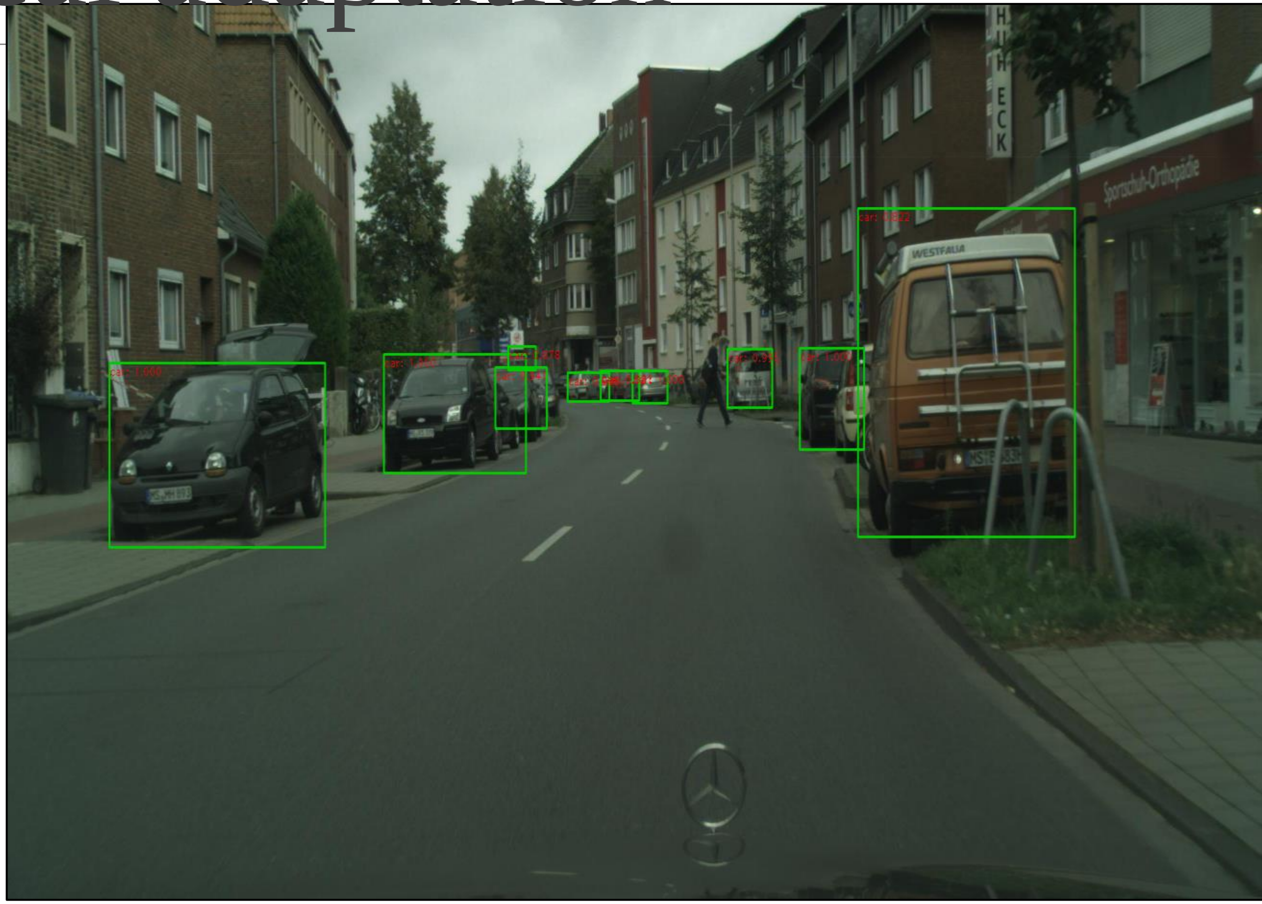
Source trained (Non-adapted)



Ours

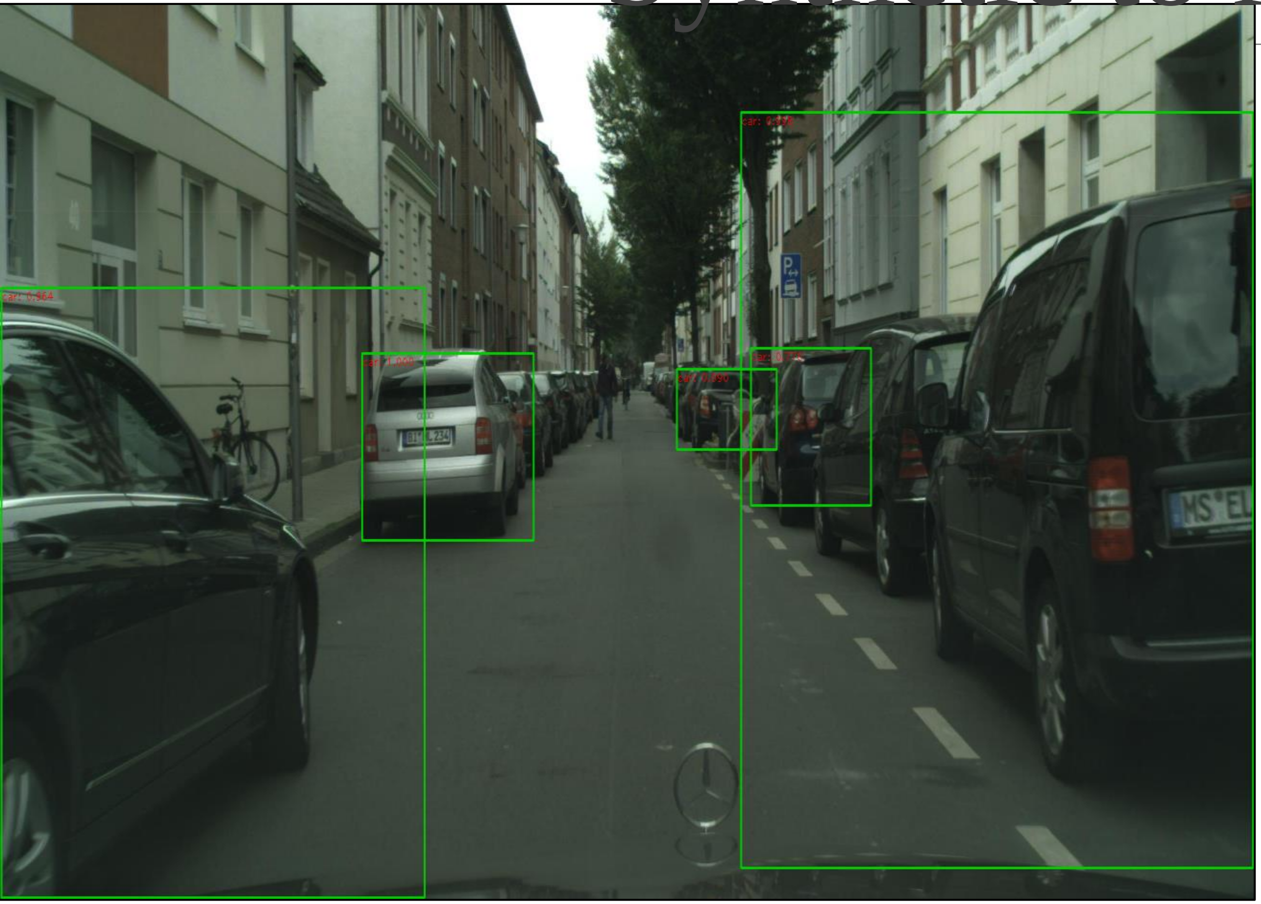
Synthetic to real adaptation

Source trained (Non-adapted)

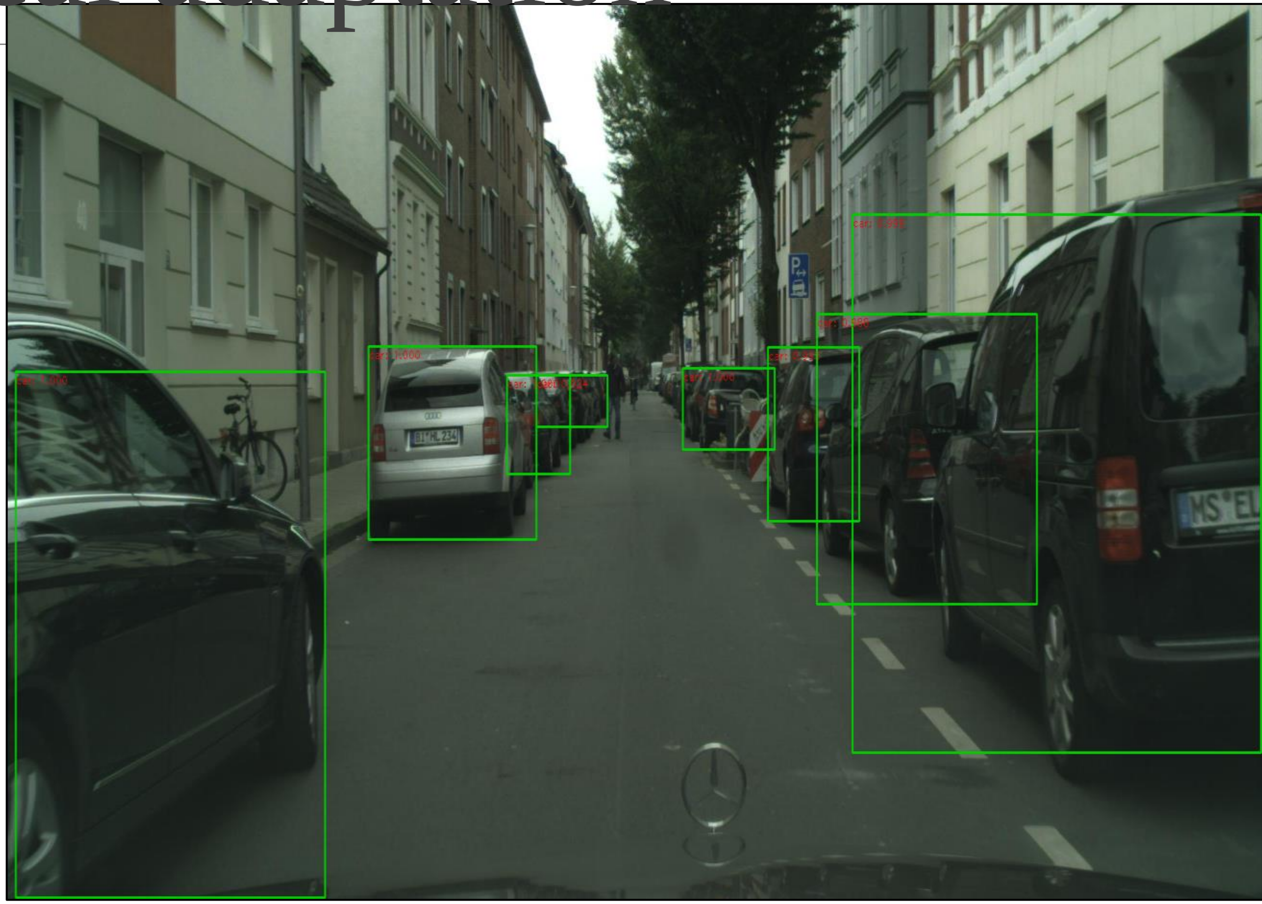


Ours

Synthetic to real adaptation

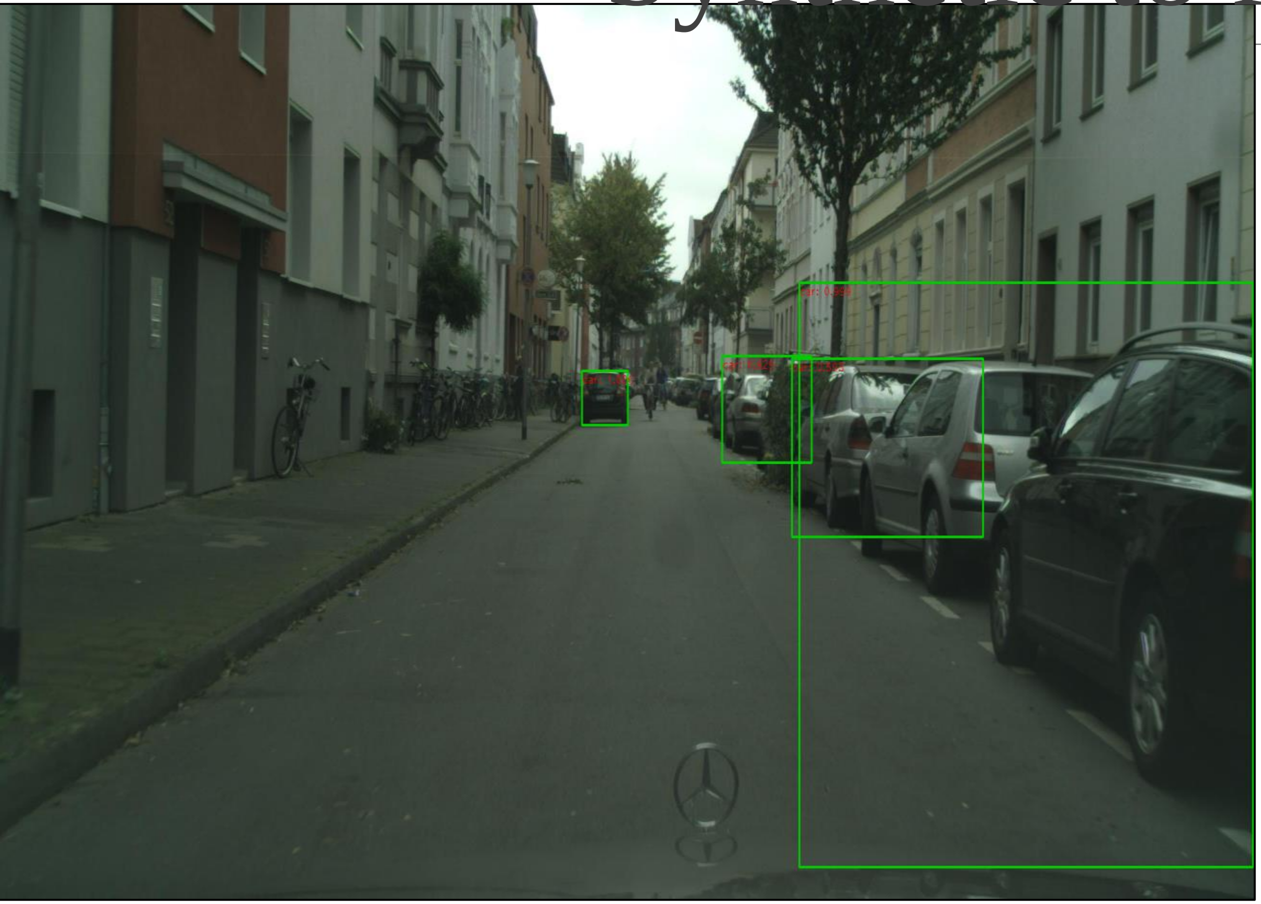


Source trained (Non-adapted)

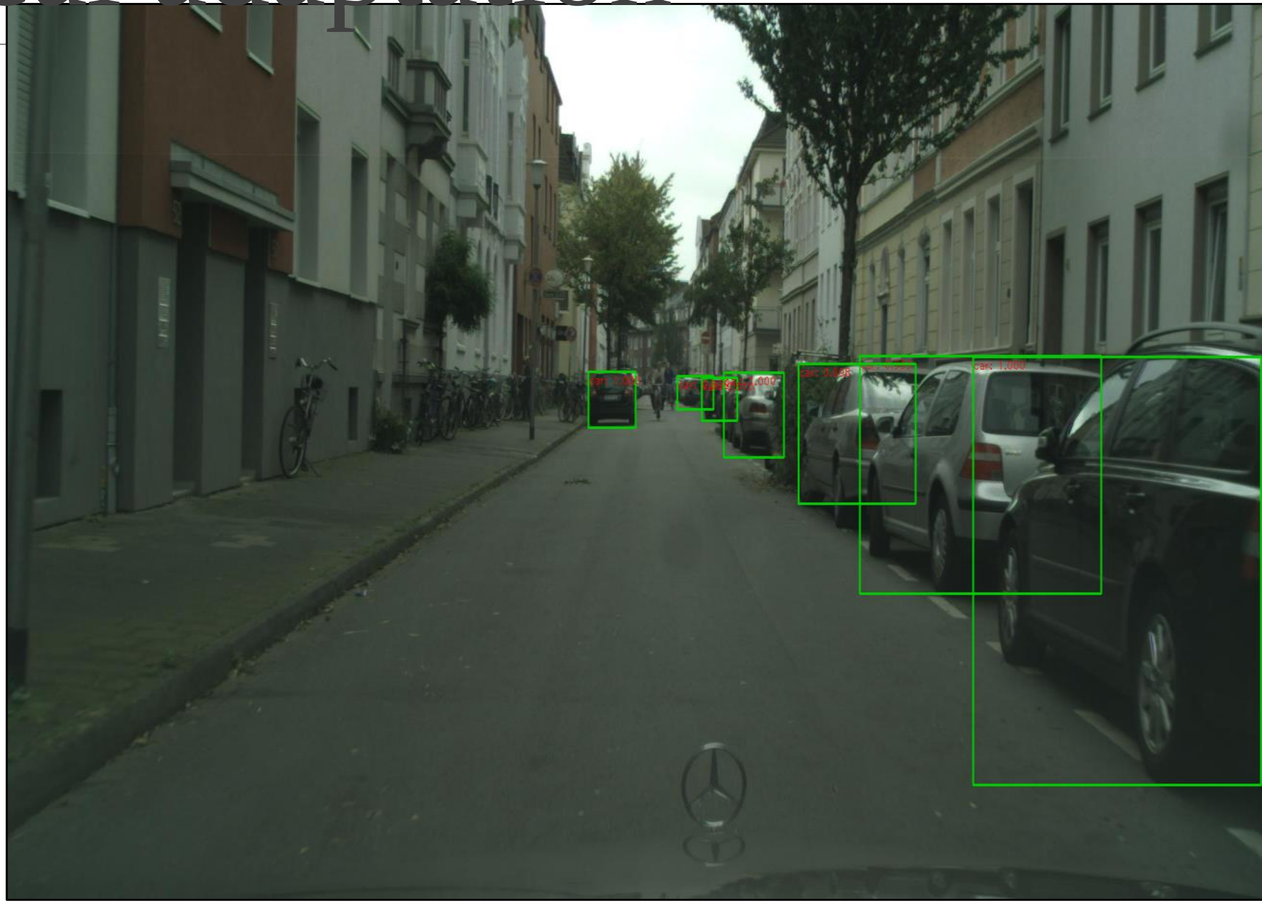


Ours

Synthetic to real adaptation

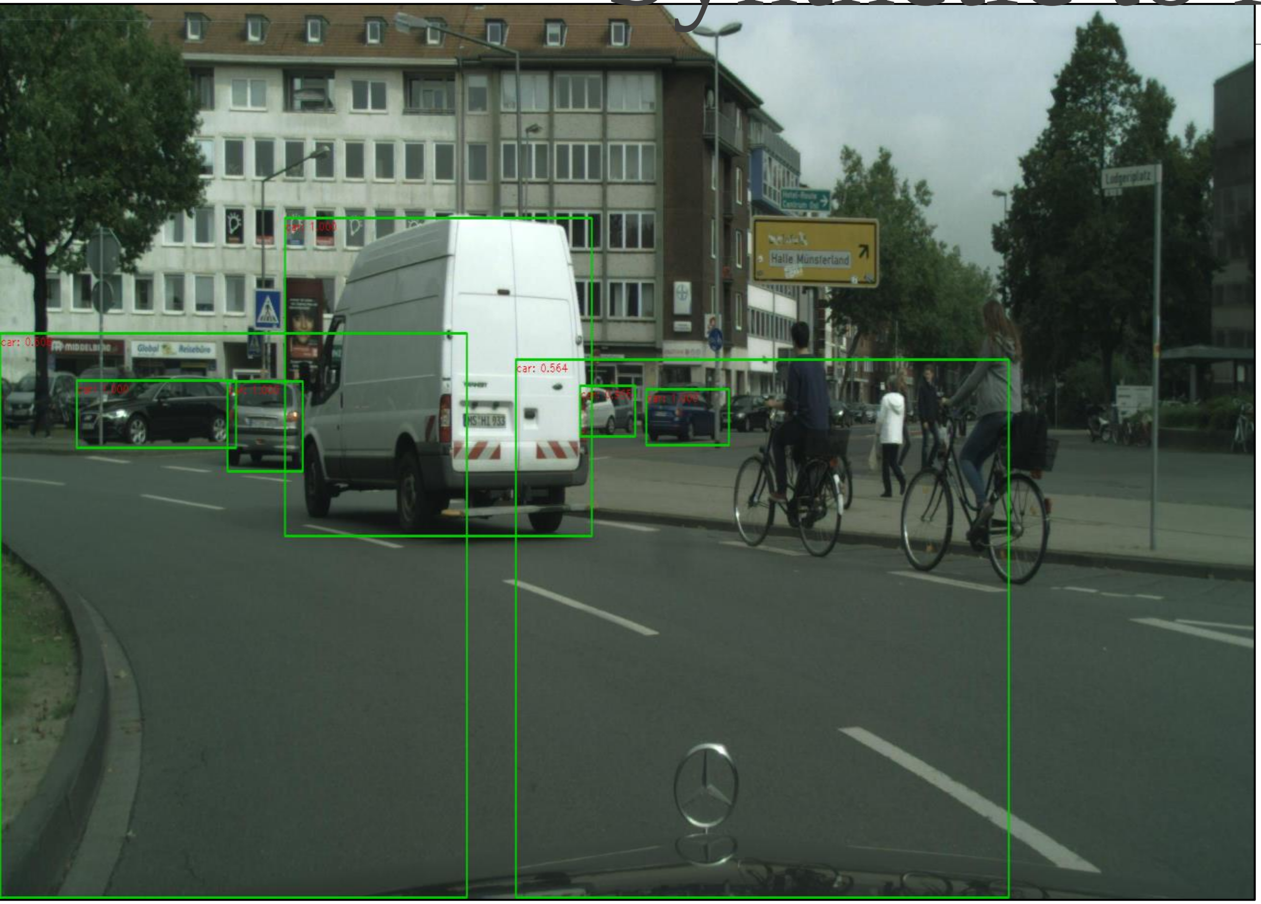


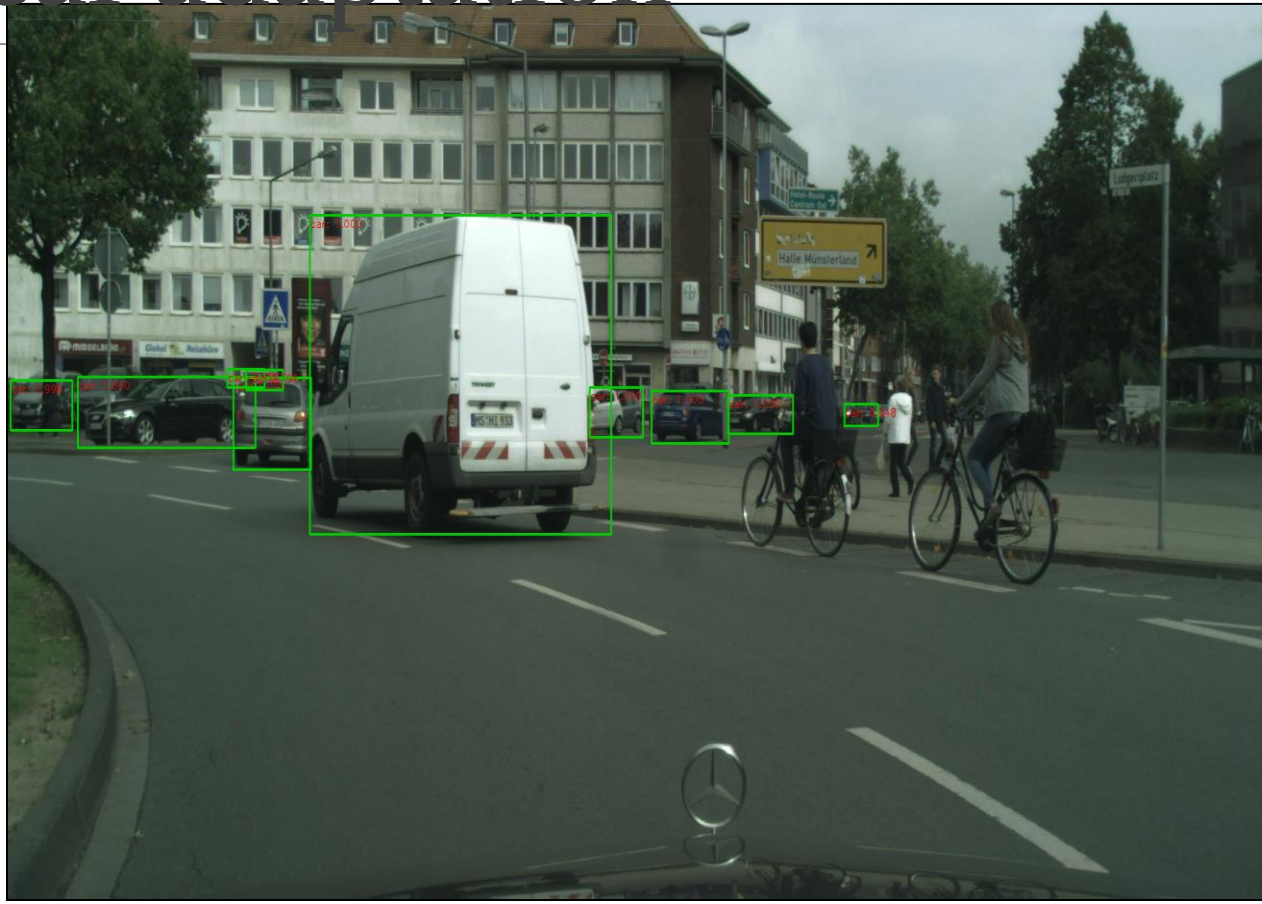
Source trained (Non-adapted)



Ours

Synthetic to real adaptation





Source trained (Non-adapted)

Ours

Adaptation results

3. CROSS-CAMERA ADAPTATION (KITTI TO CITYSCAPE)

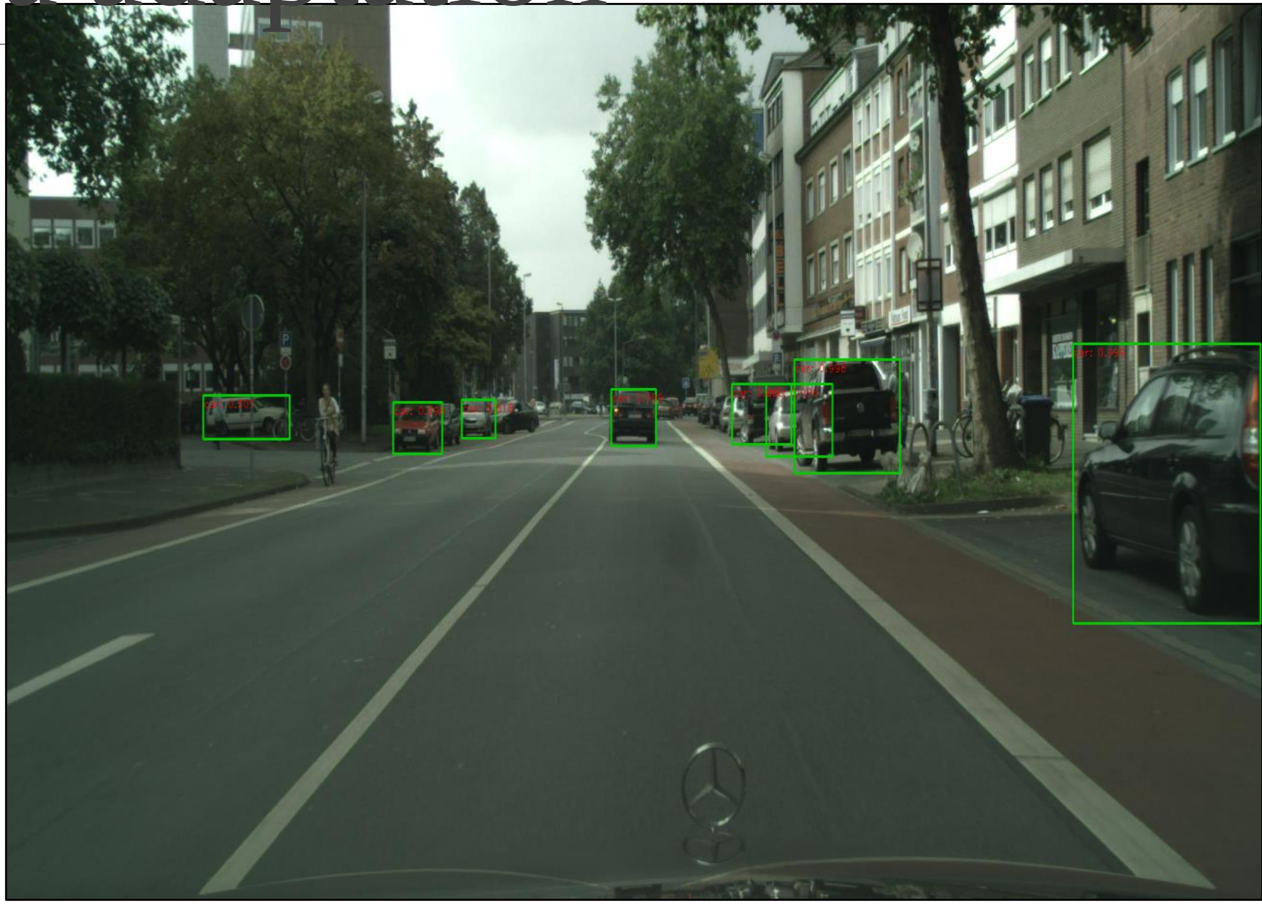
Source domain

Target domain

Adapt to

Cross-camera adaptation

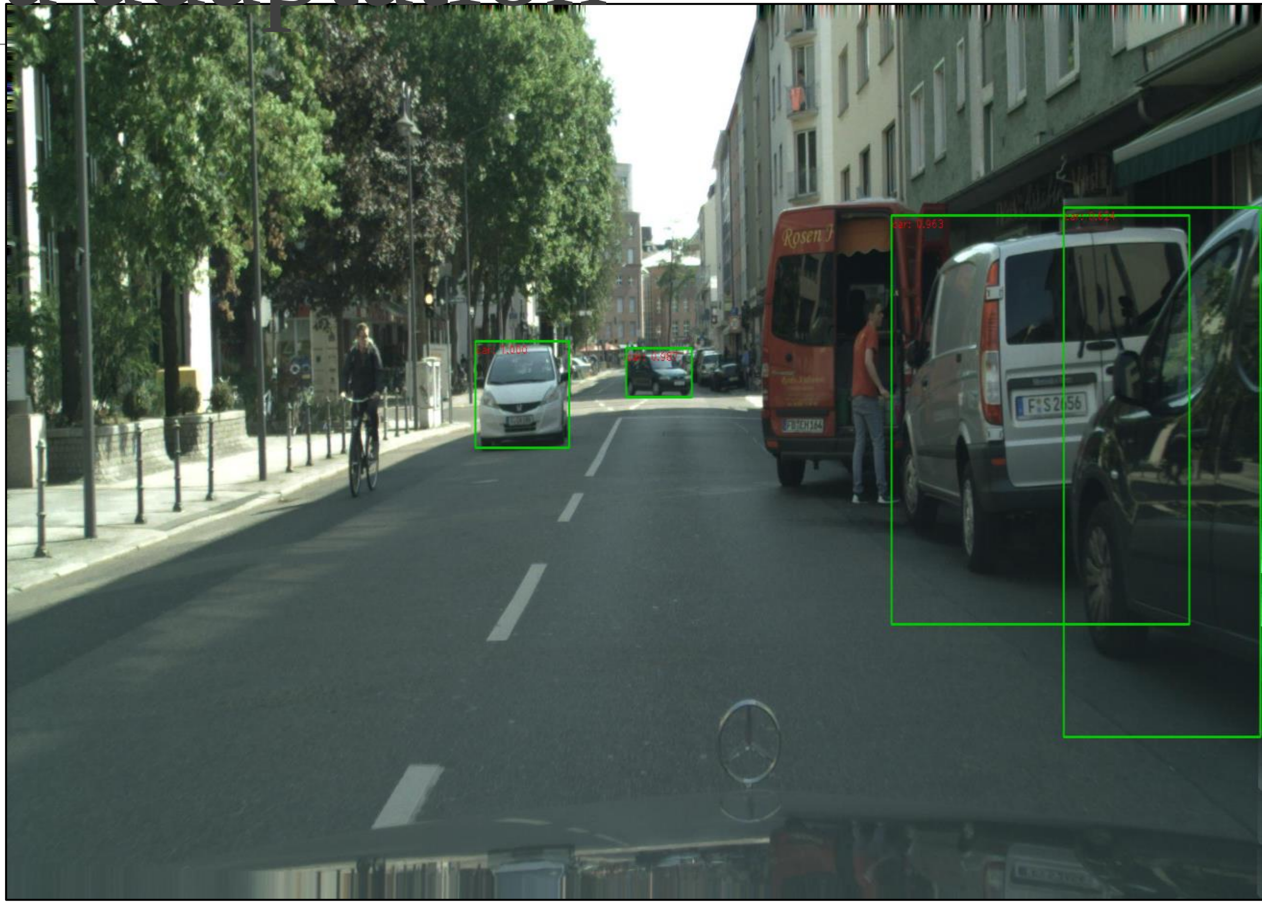
Source trained (Non-adapted)



Ours

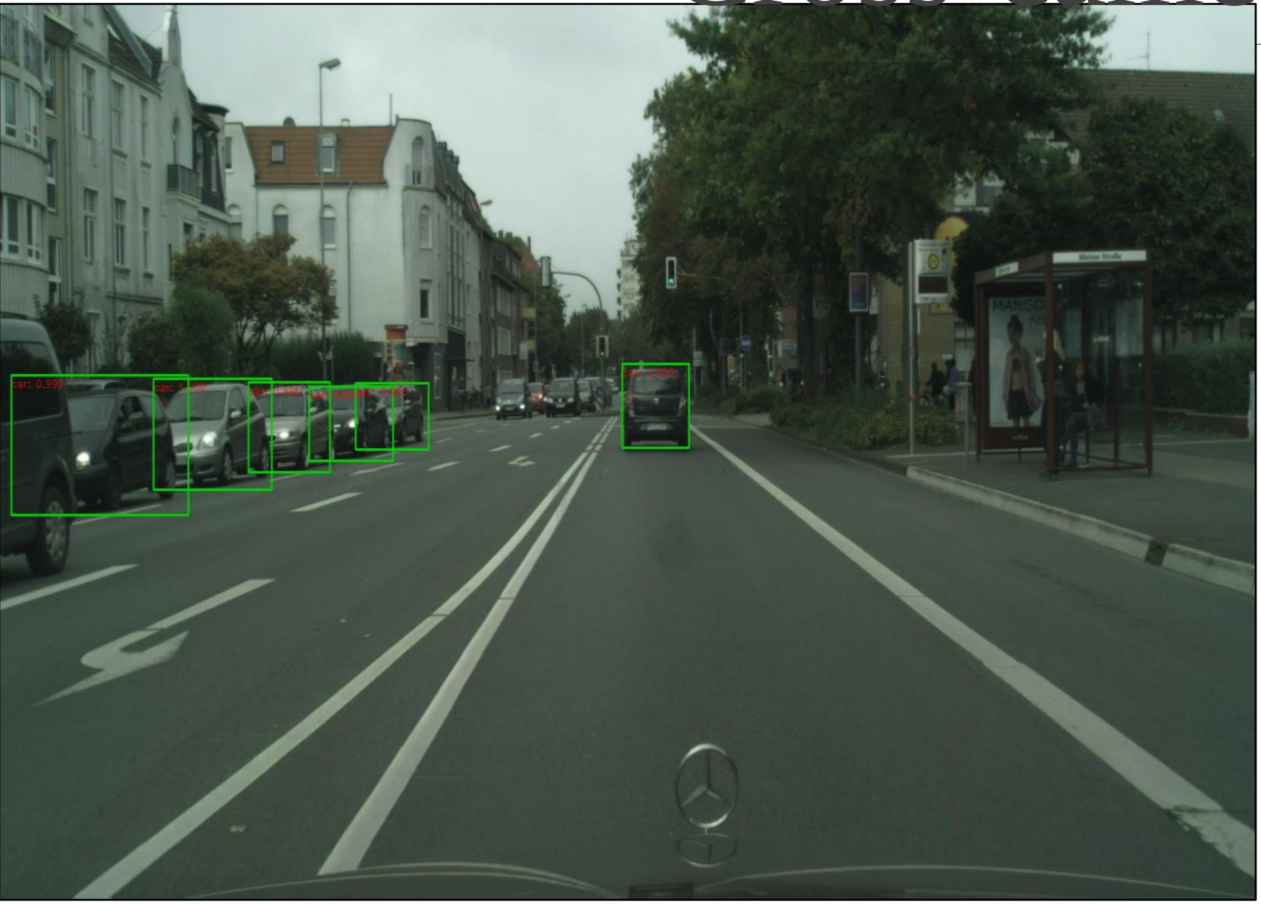
Cross-camera adaptation

Source trained (Non-adapted)

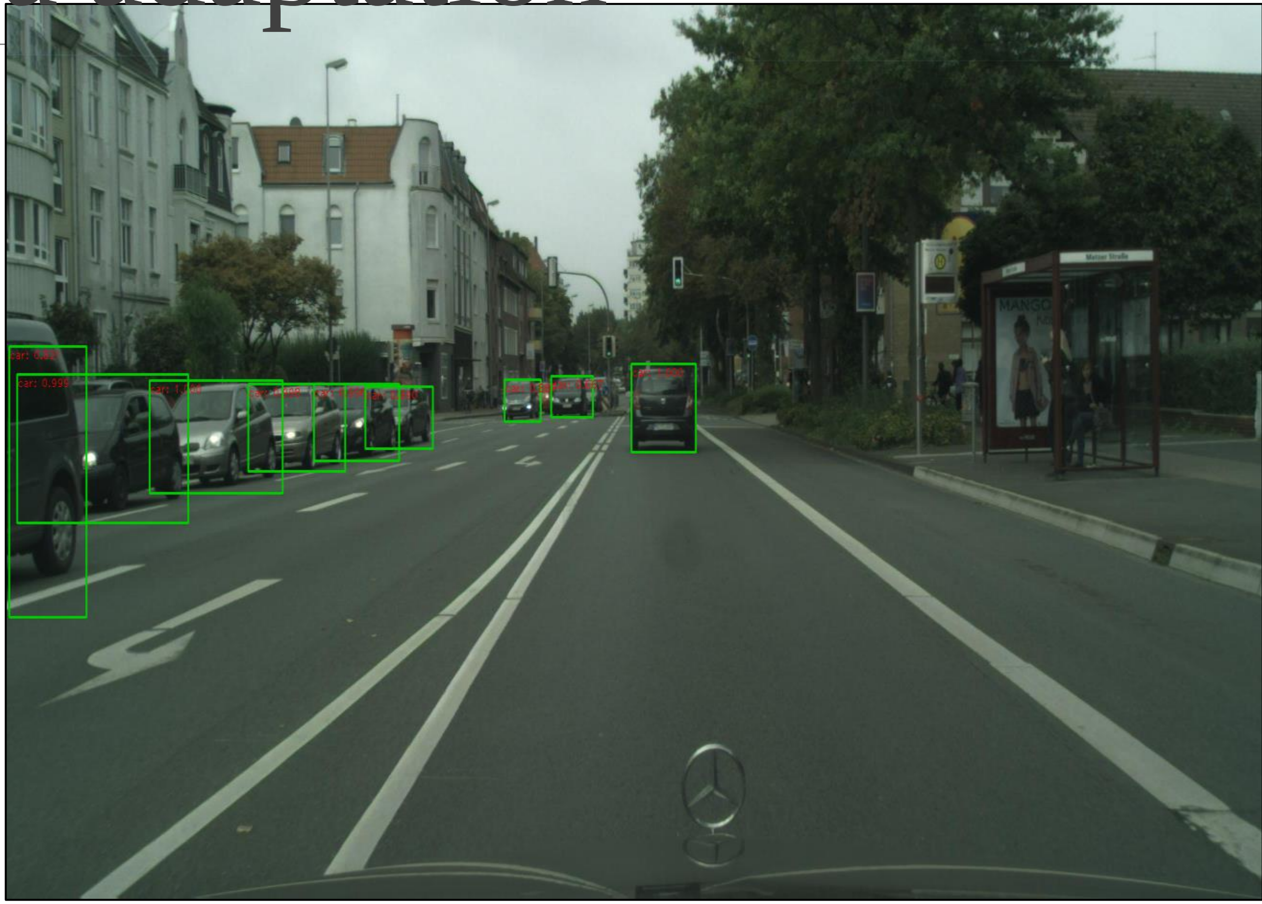


Ours

Cross-camera adaptation



Source trained (Non-adapted)



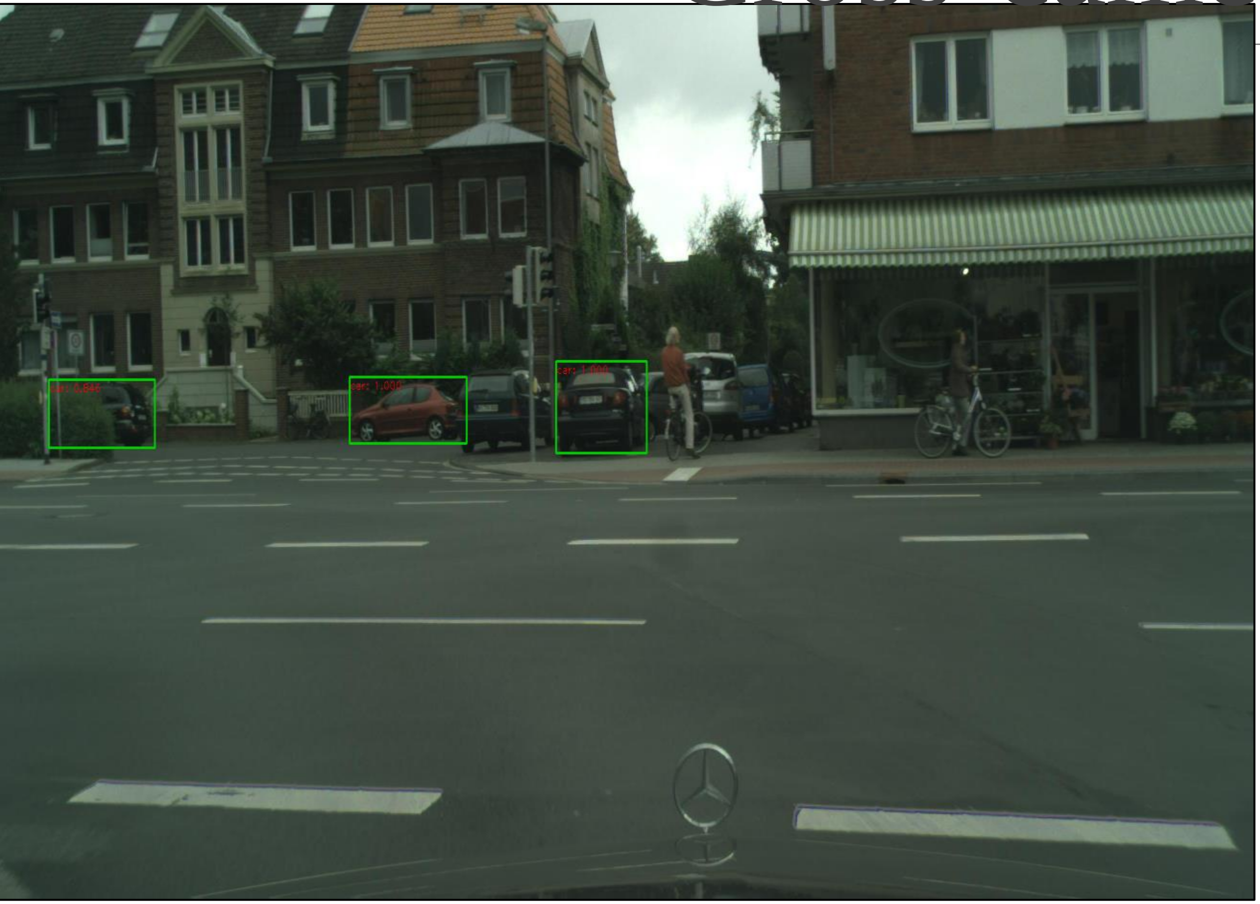
Ours

Cross-camera adaptation

Source trained (Non-adapted)

Ours

Cross-camera adaptation



Source trained (Non-adapted)

Ours