Unsupervised domain adaptation with application to urban scene analysis

Patrick Pérez

valeo.ai

Imaging and Machine Learning

4 April2019

Institut Henri Poincaré

Joint work with Tuan-Hung Vu, Himalaya Jain, Maxime Bucher and Matthieu Cord

Valeo
SMART TECHNOLOGY FOR SMARTER CARS

valeo.ai

Annotation hell

SoA visual deep learning is fully supervised

- Data collection is difficult
- Labelling is hell (if possible)

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

Annotation hell

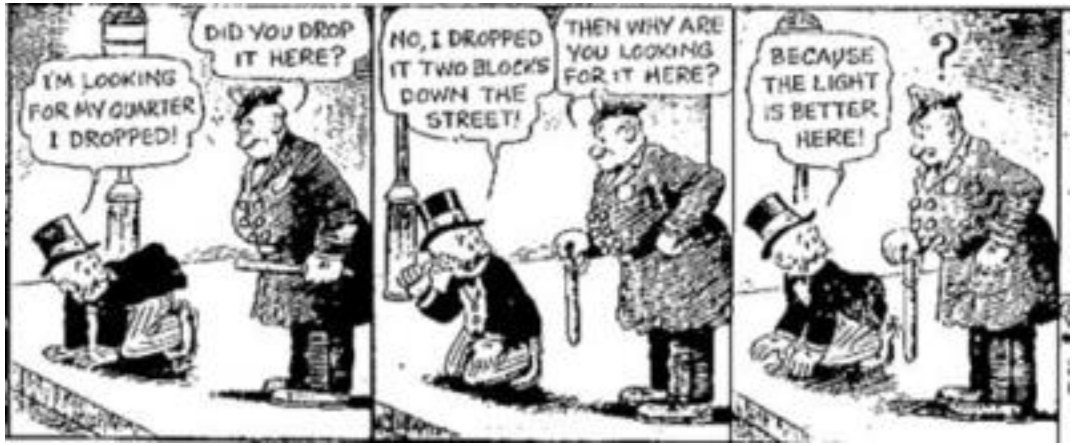
Other types of learning to the rescue

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

Transfer and adaptation

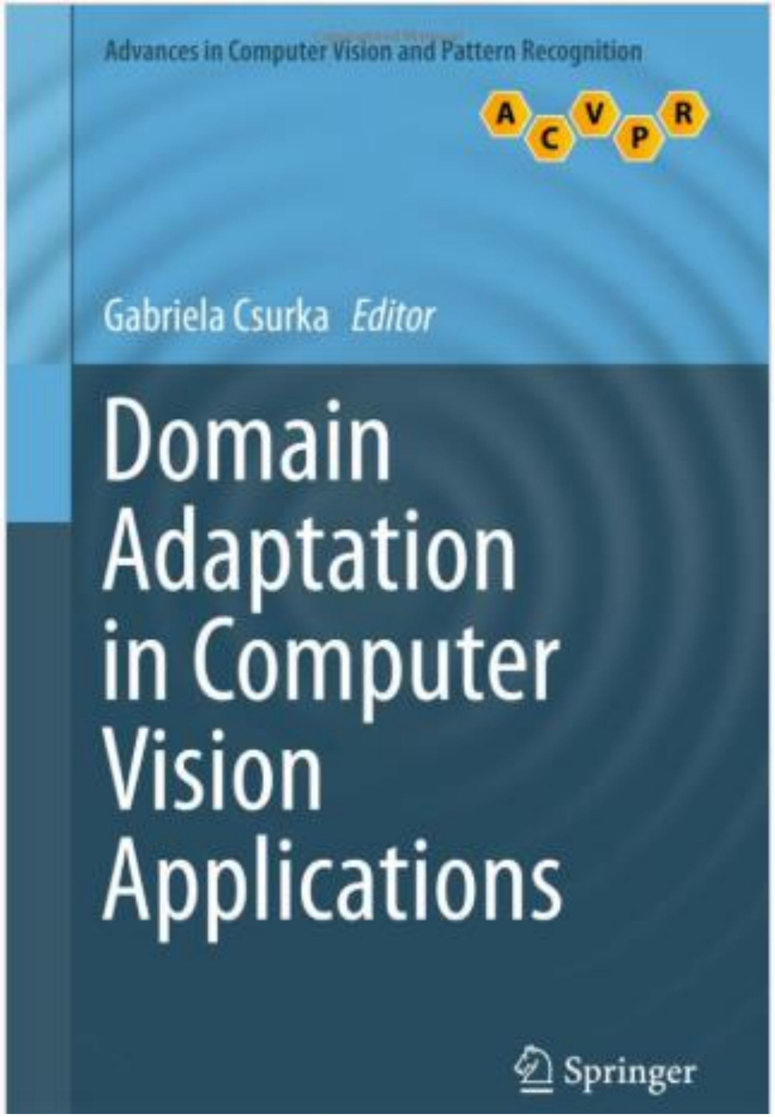
- Learn one task, conduct another one
- Learn on one distribution, run on another one = Domain Adaption

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

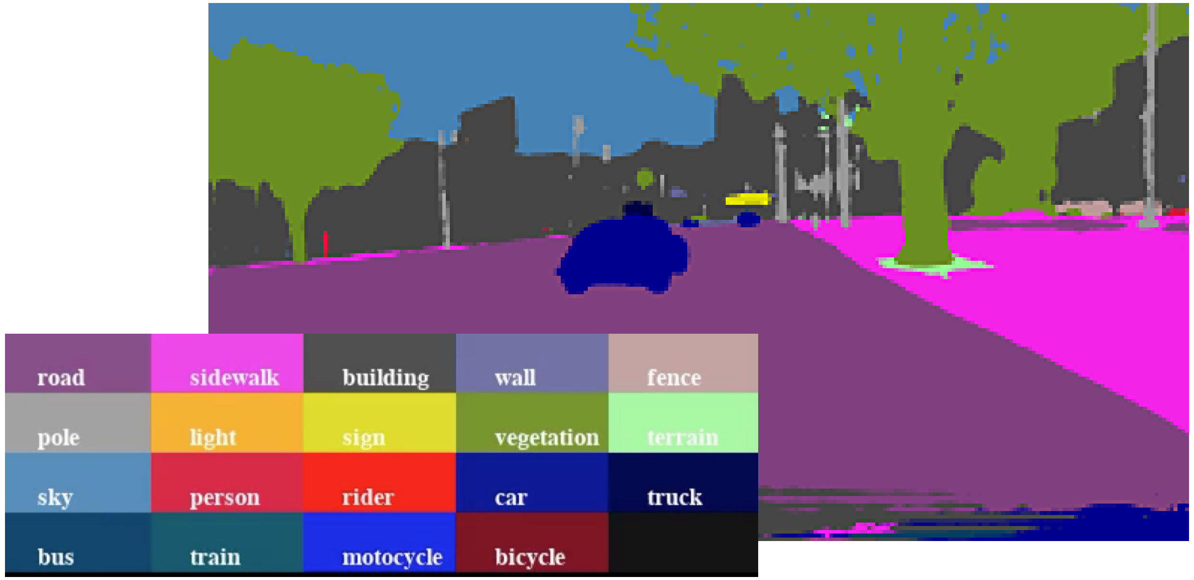
Source domain → Target domain

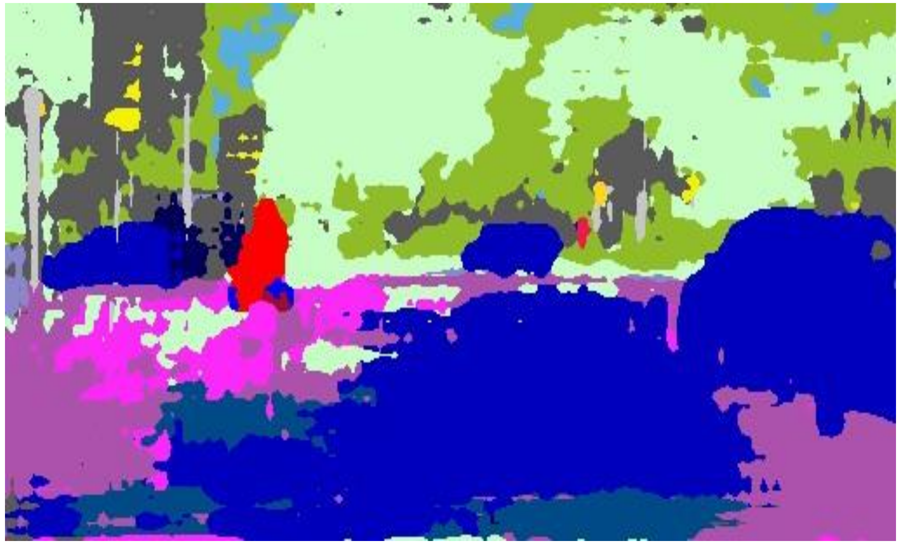
- Different weather, light, location, sensor’s spec/setup
- Synthetic vs. real

Domain gap

Different, though *related* input data distributions

Source domain → Target domain

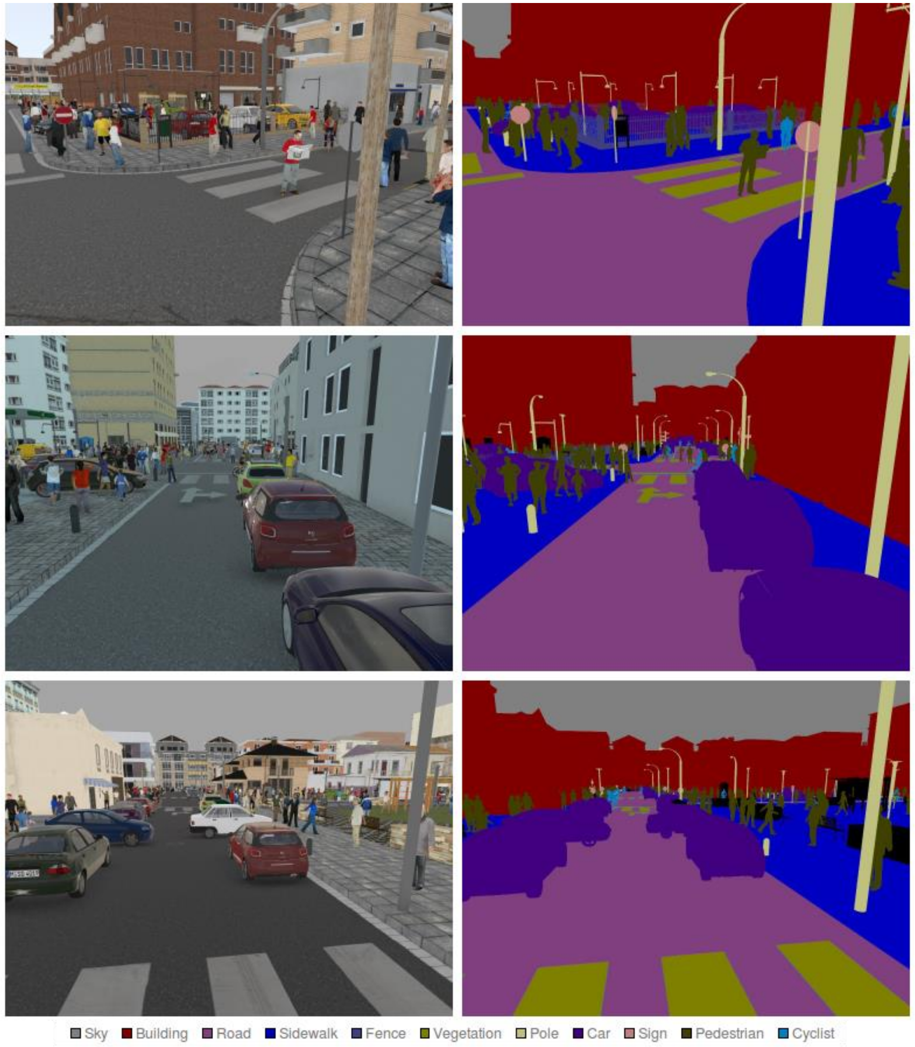




- Different weather, light, location, sensor’s spec/setup
- Synthetic vs. real

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

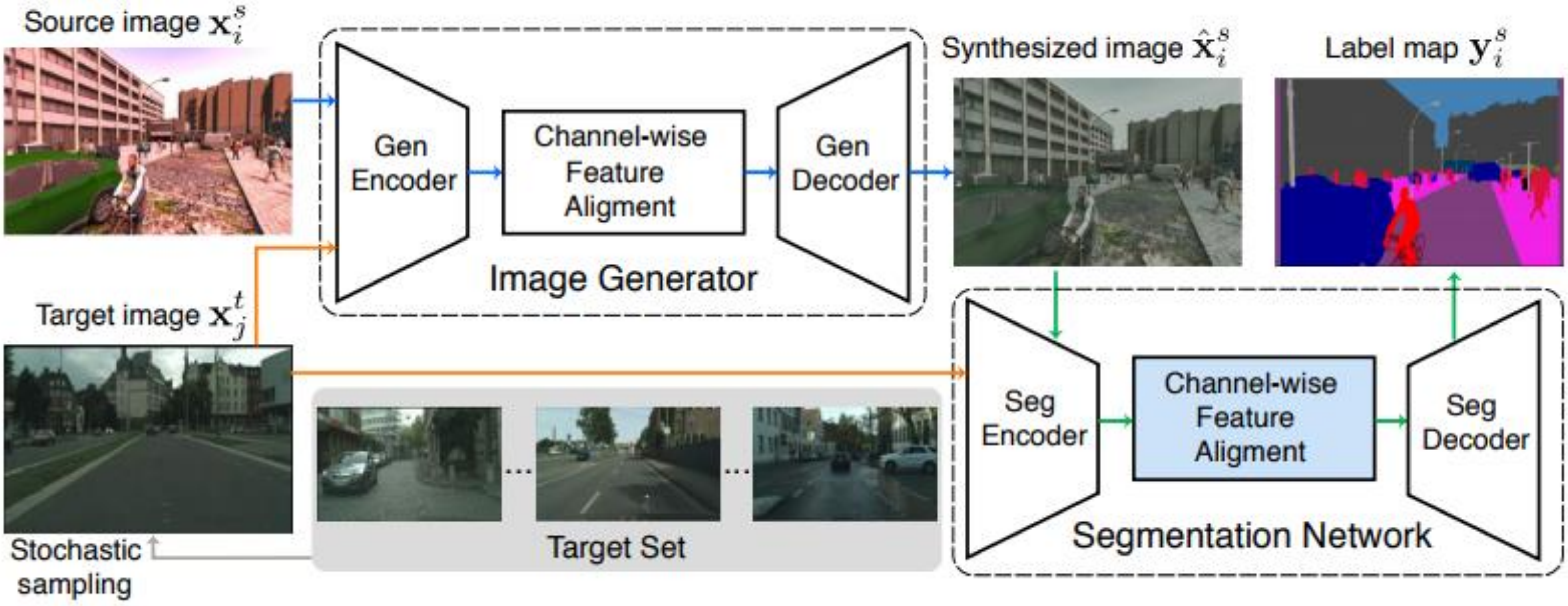
- Distribution discrepancy loss
- Optimal transport
- Discriminative adversarial loss
- Generative adversarial models

Self-training

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

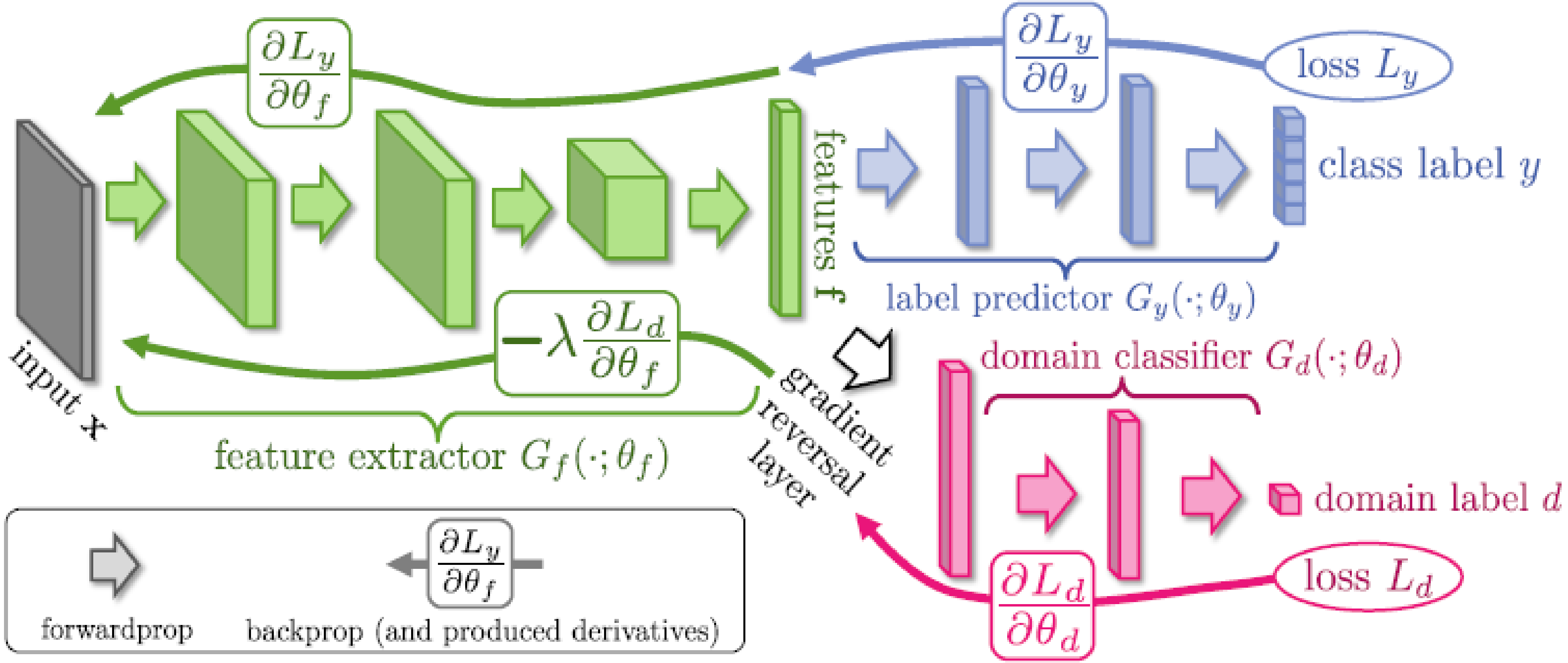
Dual Channel-wise Alignment Net (DCAN)

[Wu 2018]



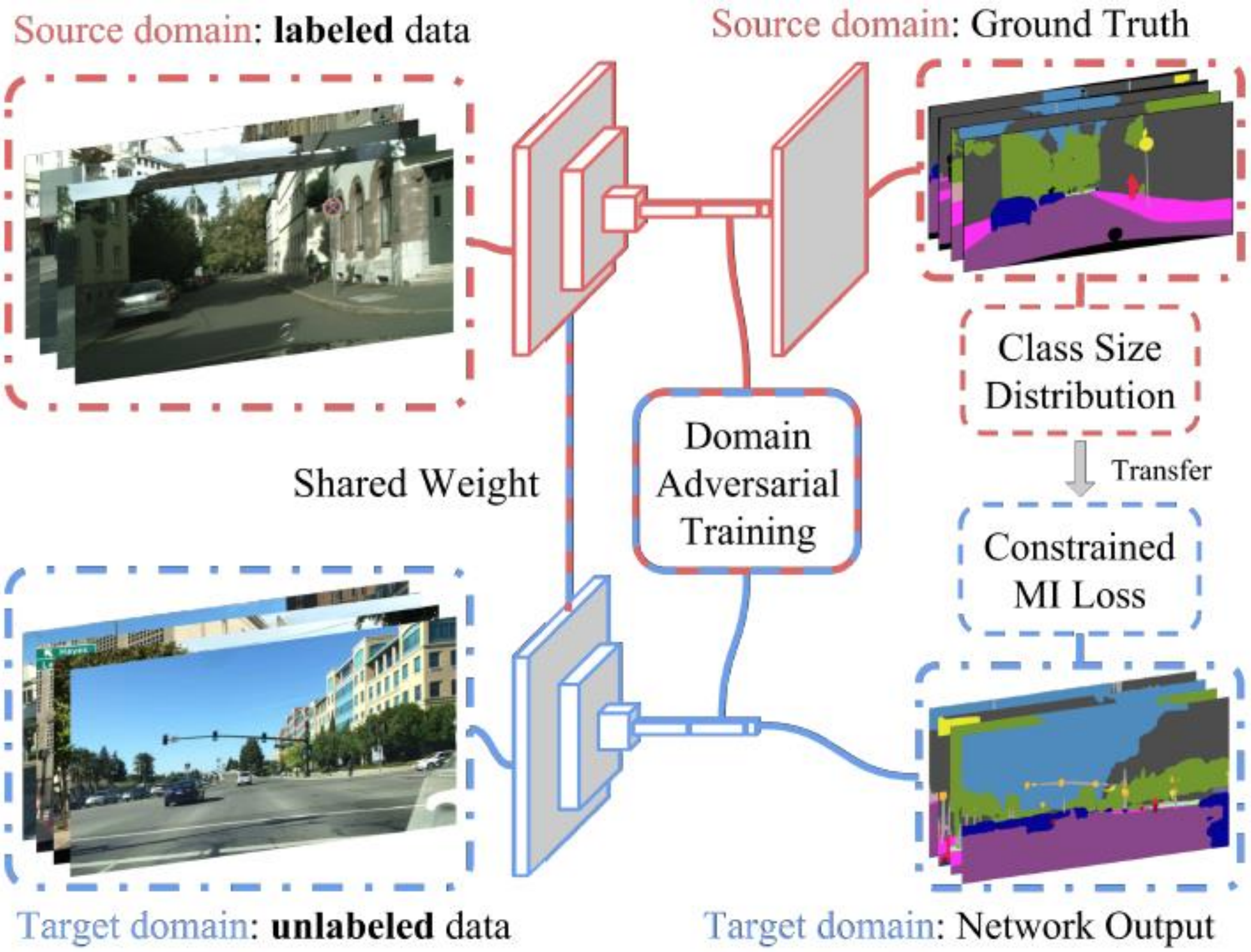
Adversarial gradient reversal

[Ganin 2015]



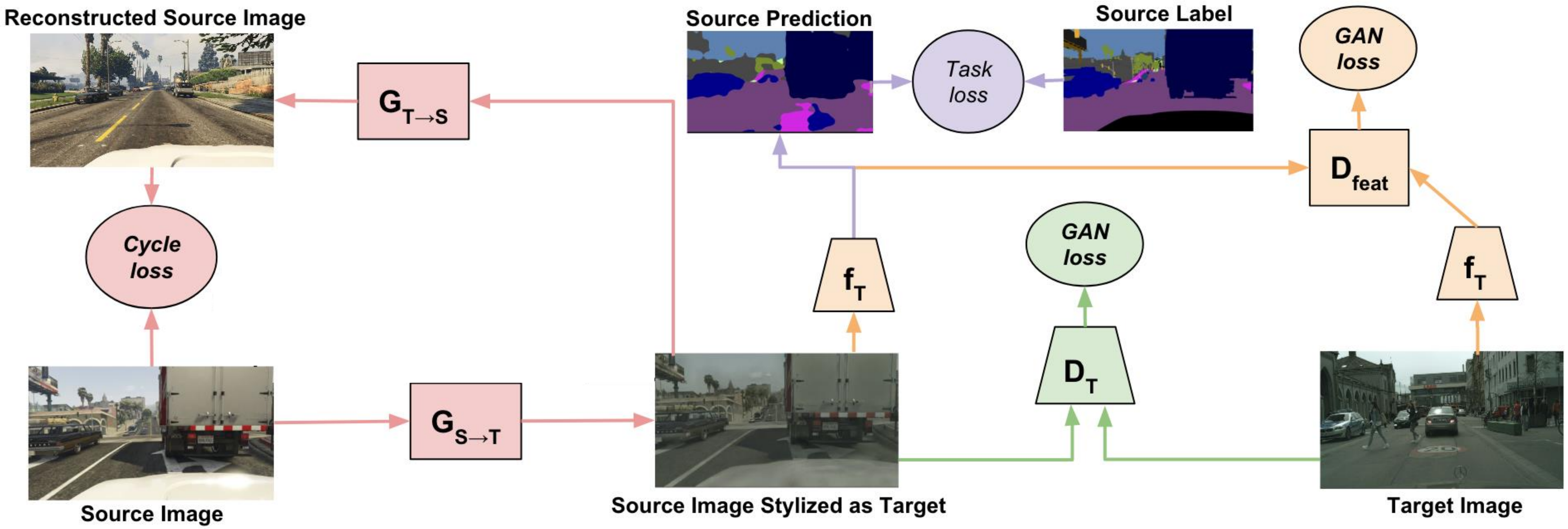
Adversarial feature alignment

[Hofmann 2016]



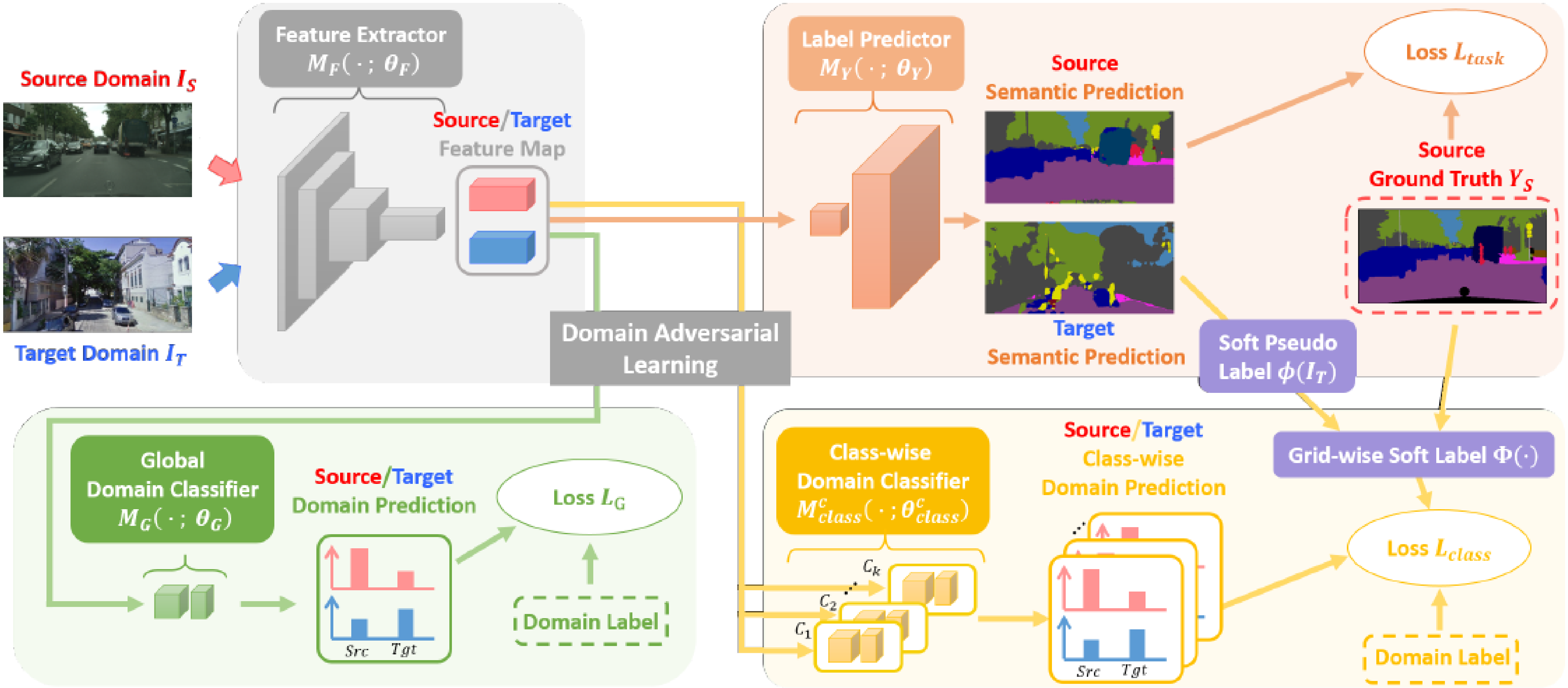
Cycle-Consistent Adversarial Domain Adaptation

CyCADA [Hoffman 2018]



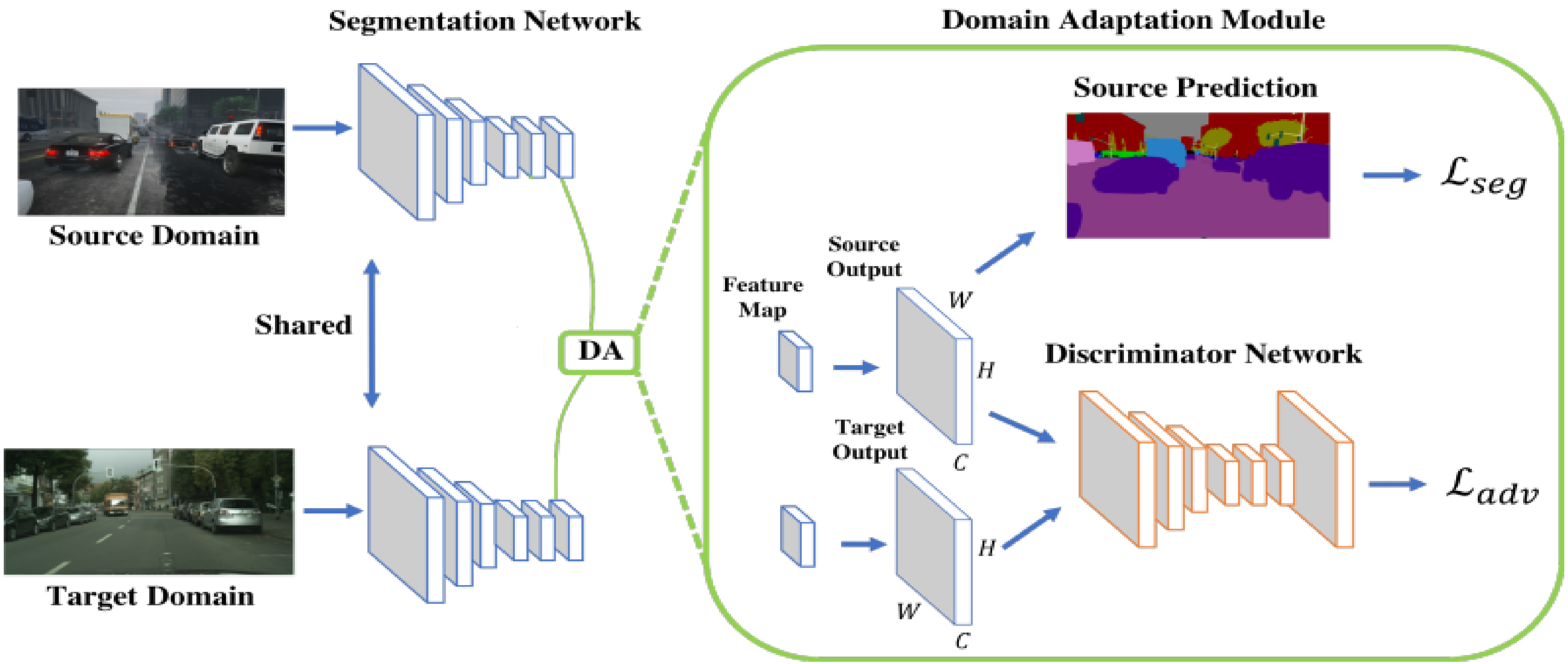
Adversarial feature alignment

[Chen 2017]



Adversarial output alignment

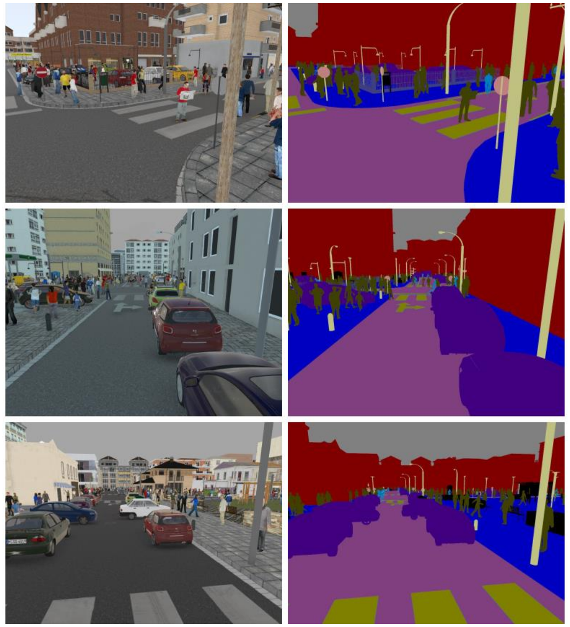
[Tsai 2018]

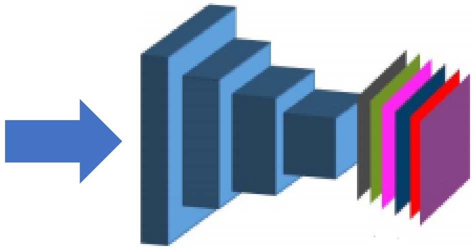


Scrutinizing output space

TRAIN

Source labelled data



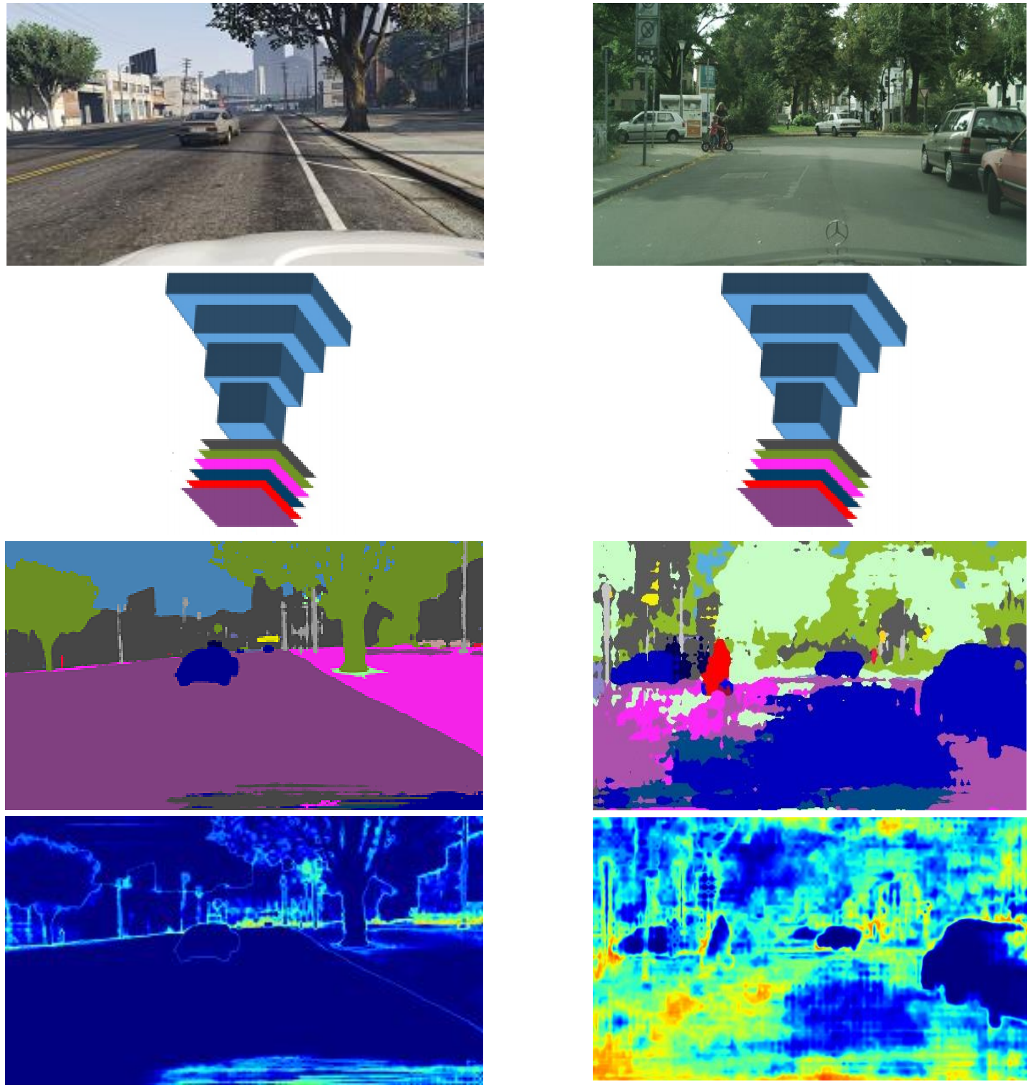


learned
segmentation
model

TEST

Source

Target



Entropy-based alignments

Observations

- Entropy way higher/noisier on target vs. source inputs
- Entropy is class-agnostic and requires no label to be computed

Proposed approach [Vu 2019]

- Pixel-wise minimization of predictive entropy on target (*MinEnt*)
- Image-wise adversarial discrimination of information maps (*AdvEnt*)

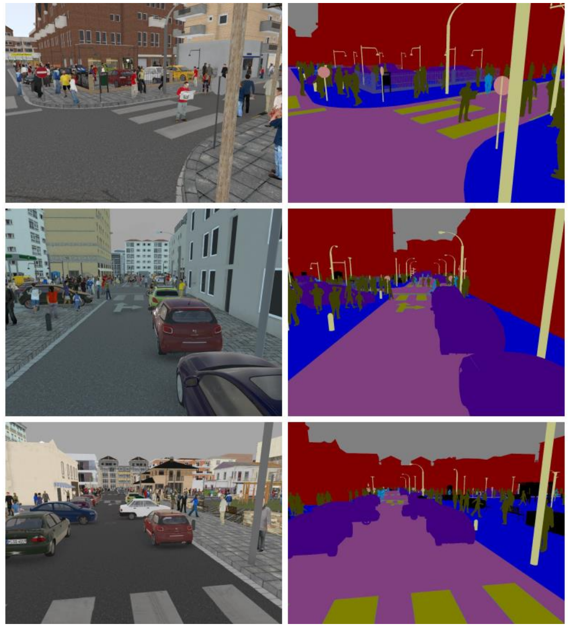
Aim

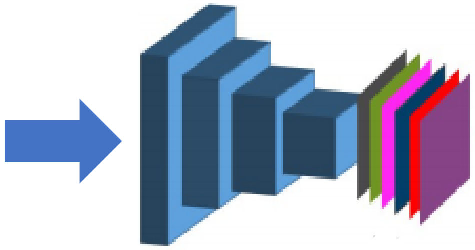
- Increase prediction confidence
- Reduce domain gap

Scrutinizing output space

TRAIN

Source labelled data

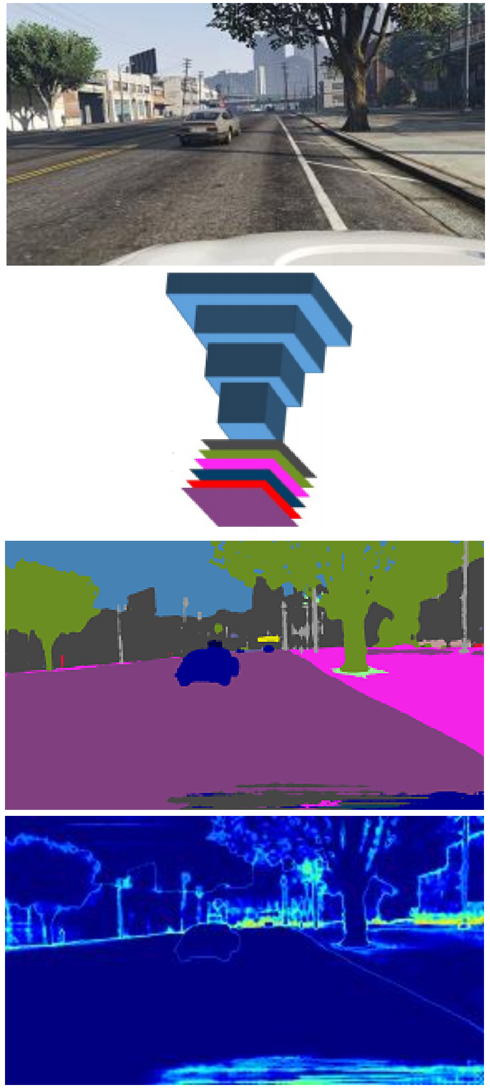




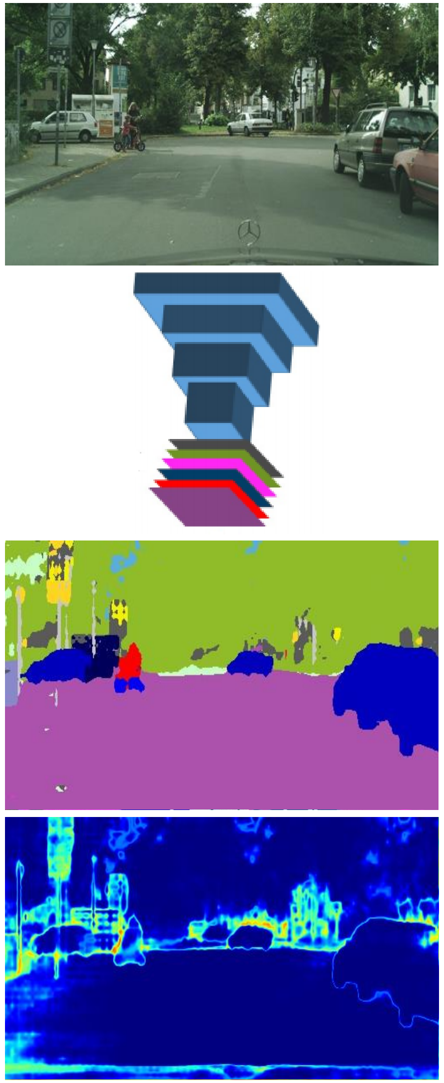
learned segmentation model

TEST

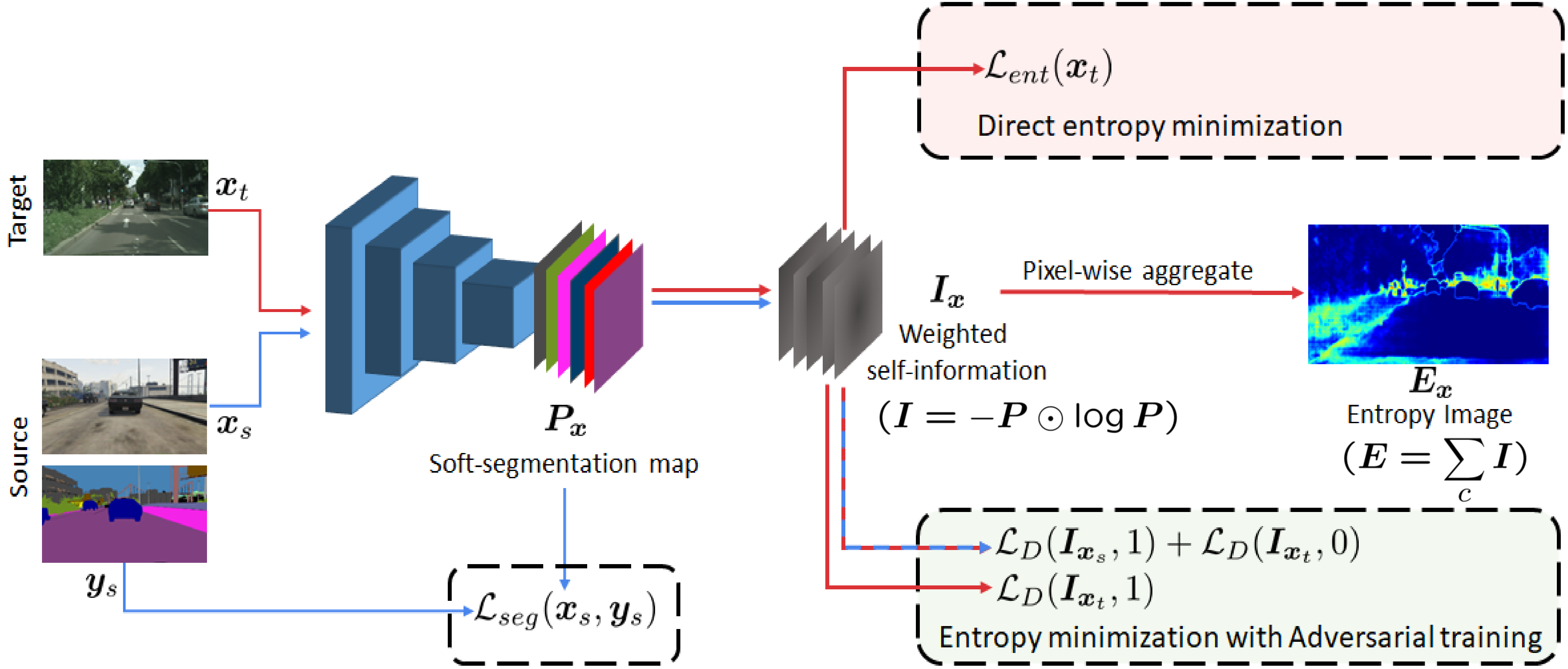
Source



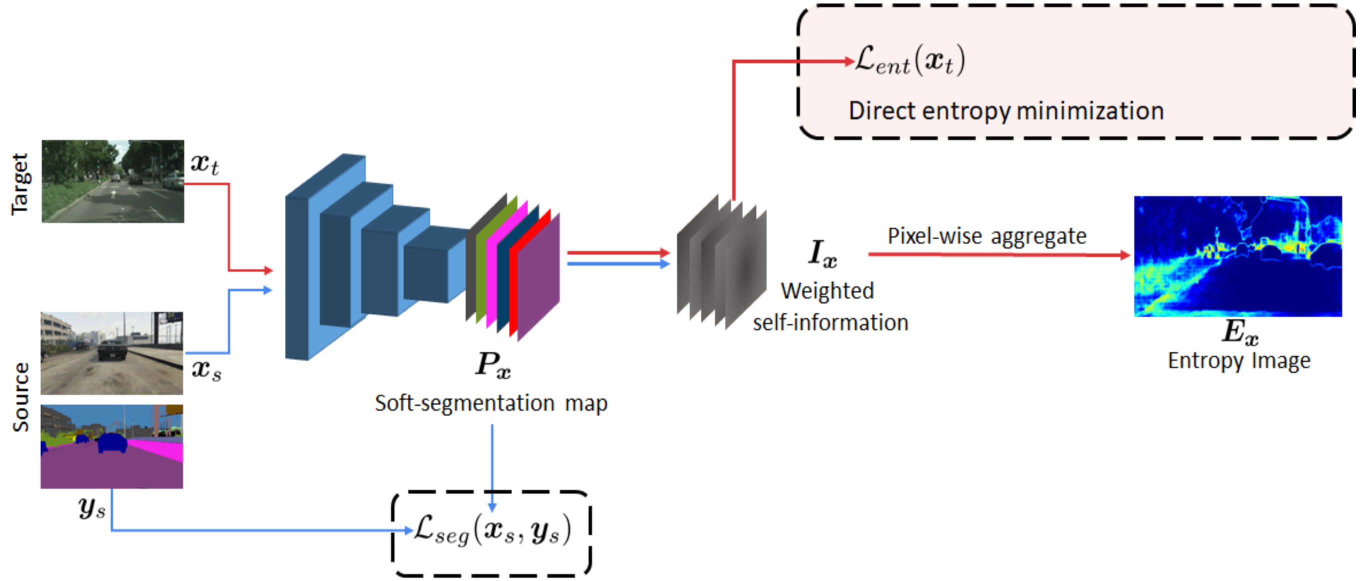
Target



Proposed method



Proposed method



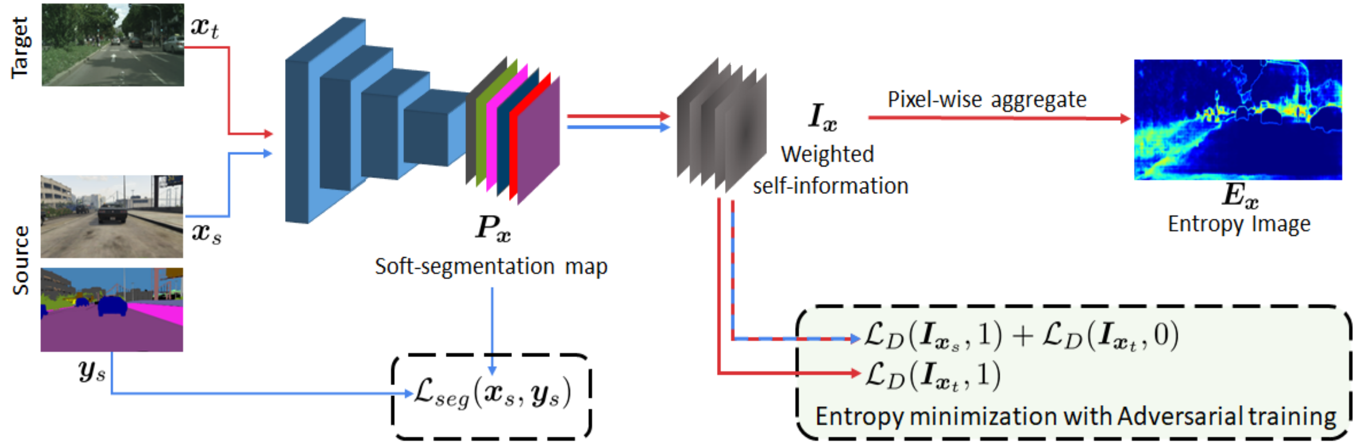
$$\mathcal{L}_{ent}(\boldsymbol{x}_t) = \sum_{h,w} \boldsymbol{E}_{\boldsymbol{x}_t}^{(h,w)}$$

$$\mathcal{L}_{seg}(\boldsymbol{x}_s, \boldsymbol{y}_s) = -\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C} \boldsymbol{y}_s^{(h,w,c)} \log \boldsymbol{P}_{\boldsymbol{x}_s}^{(h,w,c)}$$

MinEnt

$$\min_{\theta_F} \frac{1}{|\mathcal{X}_s|} \sum_{\boldsymbol{x}_s} \mathcal{L}_{seg}(\boldsymbol{x}_s, \boldsymbol{y}_s) + \frac{\lambda_{ent}}{|\mathcal{X}_t|} \sum_{\boldsymbol{x}_t} \mathcal{L}_{ent}(\boldsymbol{x}_t)$$

Proposed method



$$\mathcal{L}_{seg}(\boldsymbol{x}_s, \boldsymbol{y}_s) = -\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C} \boldsymbol{y}_s^{(h,w,c)} \log \boldsymbol{P}_{\boldsymbol{x}_s}^{(h,w,c)}$$

AdvEnt

$$\min_{\theta_D} \frac{1}{|\mathcal{X}_s|}\sum_{\boldsymbol{x}_s} \mathcal{L}_D(\boldsymbol{I}_{\boldsymbol{x}_s}, 1) + \frac{1}{|\mathcal{X}_t|}\sum_{\boldsymbol{x}_t} \mathcal{L}_D(\boldsymbol{I}_{\boldsymbol{x}_t}, 0)$$

$$\min_{\theta_F} \frac{1}{|\mathcal{X}_s|}\sum_{\boldsymbol{x}_s} \mathcal{L}_{seg}(\boldsymbol{x}_s, \boldsymbol{y}_s) + \frac{\lambda_{adv}}{|\mathcal{X}_t|}\sum_{\boldsymbol{x}_t} \mathcal{L}_D(\boldsymbol{I}_{\boldsymbol{x}_t}, 1)$$

Experiments

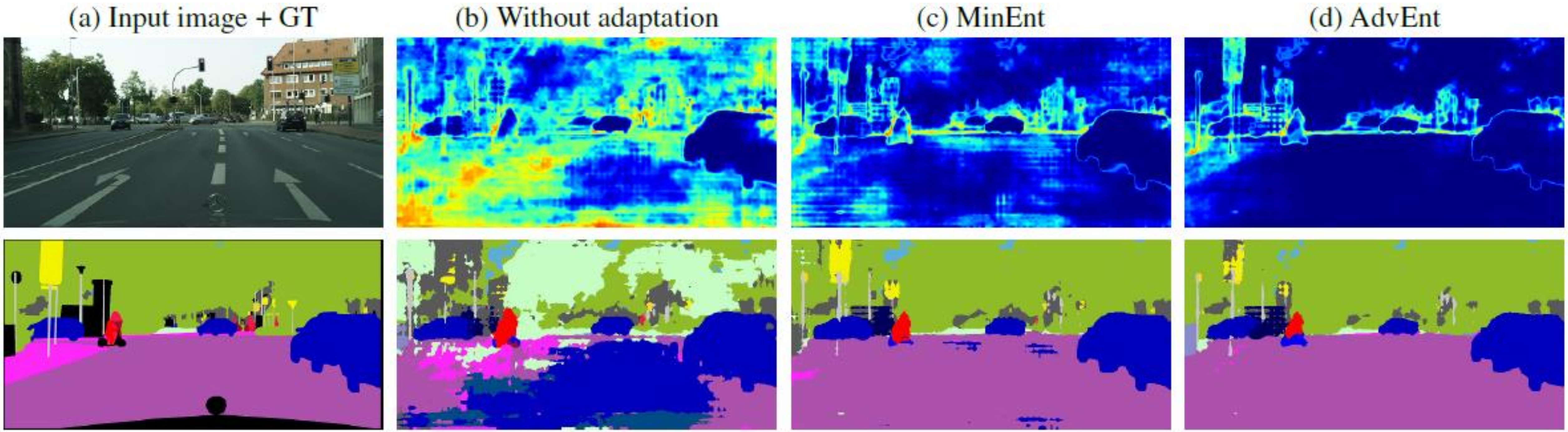
Ingredients

- Segmentor: DeepLab2 (with VGG-16 or ResNet-101)
- Discriminator: as in DCGAN (4 conv-layers, leaky ReLU)
- Metric: mean Intersection over Union (mIoU %)

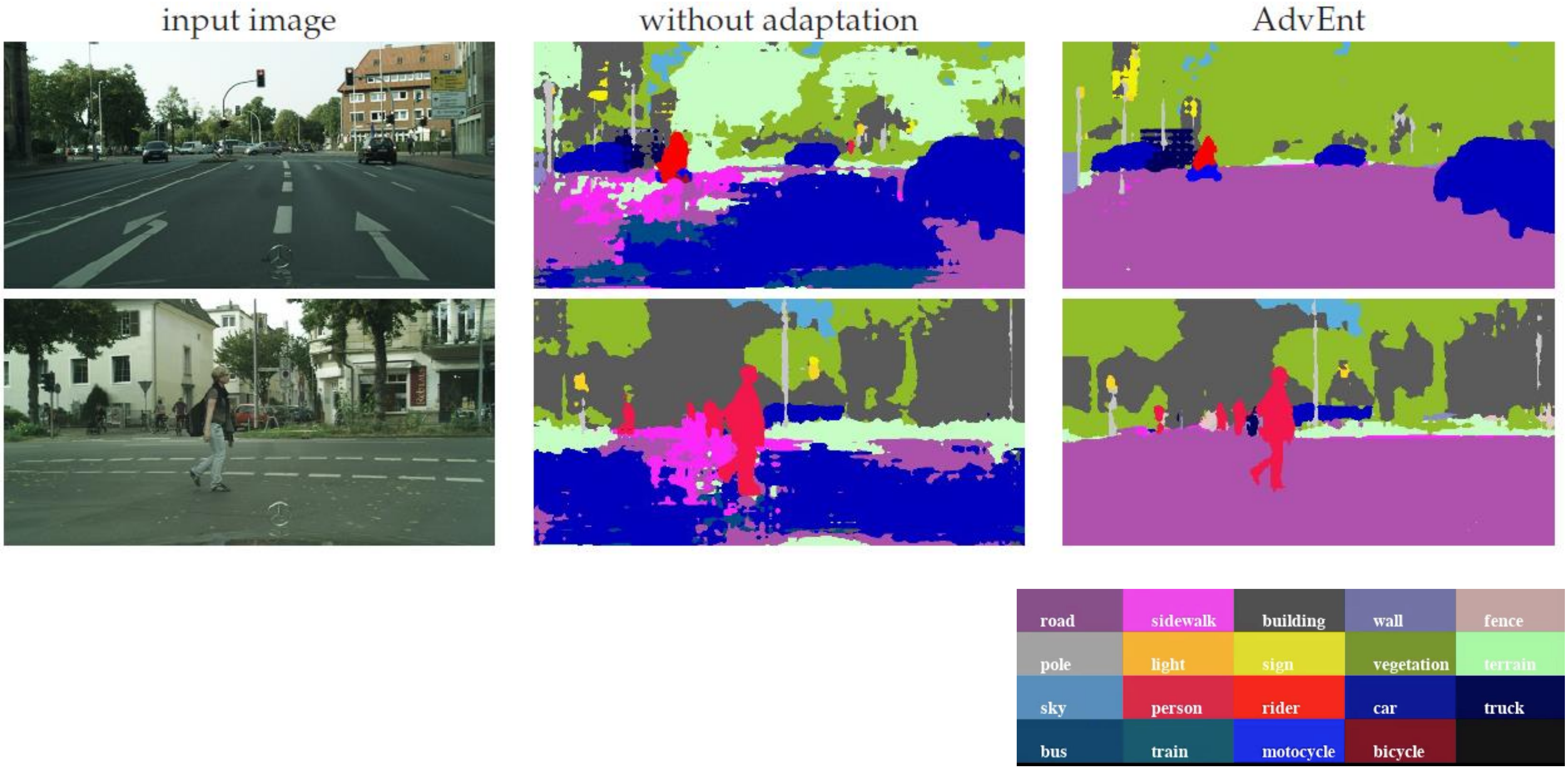
Synthetic-to-Real setups

- GTA5→Cityscapes: 25k/3k images, 19 common classes
- Synthia→Cityscapes: 10k/3k images, 16 common classes

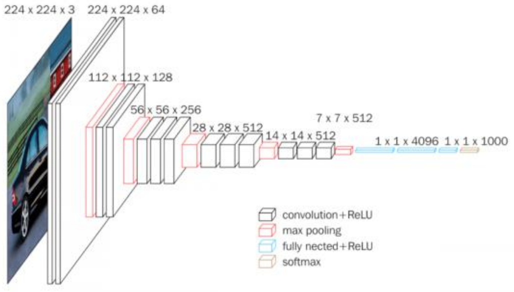
Qualitative results

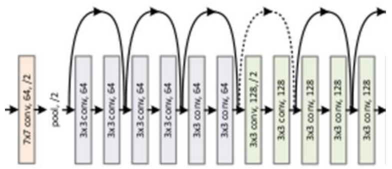


Qualitative results



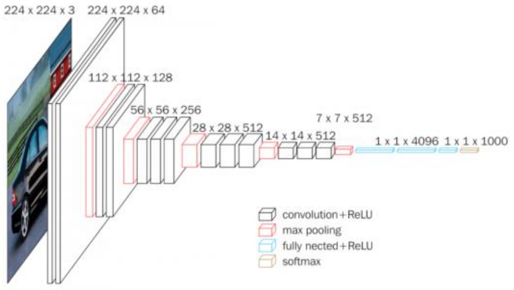
Quantitative results

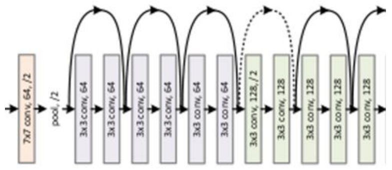




GTA5 → Cityscapes

Method	UDA Model	Oracle	mIoU Gap (%)
FCNs [Hoffman'16]	27.1	64.6	-37.5
CyCADA [Hoffman'18]	28.9	60.3	-31.4
Adapt-SegNet [Tsai'18]	35.0	61.8	-25.2
Ours (single model)	35.6	61.8	-24.6
Adapt-SegNet [Tsai'18]	42.4	65.1	-22.7
Ours (single model)	43.8	65.1	-21.3
Ours (two models)	45.5	65.1	**-19.6**





SYNTHIA → Cityscapes

Method	UDA Model	Oracle	mIoU Gap (%)
FCNs [Hoffman'16]	22.9	73.8	-50.9
Adapt-SegNet [Tsai'18]	37.6	68.4	-30.8
Ours (single model)	36.6	68.4	-31.8
Adapt-SegNet [Tsai'18]	46.7	71.7	-25.0
Ours (single model)	47.6	71.7	-24.1
Ours (two models)	48	71.7	**-23.7**

Extension to object detection

Cityscapes → Cityscapes Foggy

Models	person	rider	car	truck	bus	train	mcycle	bicycle	mAP
SSD-300	15.0	17.4	27.2	5.7	15.1	9.1	11.0	16.7	14.7
Ours (MinEnt)	15.8	22.0	28.3	5.0	15.2	15.0	13.0	20.6	16.9
Ours (AdvEnt)	**17.6**	**25.0**	**39.6**	**20.0**	**37.1**	**25.9**	**21.3**	**23.1**	**26.2**

Outlook

Important problem, big improvement margin

Hybridization with

- Multi-sensor, multi-task
- Other orthogonal methods (‘’conservative loss’’, class prior)
- Privileged information learning on source domain
- Semi-supervised/few-shot learning
- Learning w/o forgetting (toward “domain extension”)

References

[Wu 2018] Z. Wu, X. Han, Y.-L. Lin, M. Gokhan Uzunbas, T. Goldstein, S. Nam Lim and L. S. Davis.
DCAN: Dual channel-wise alignment networks for unsupervised scene adaptation. ECCV 2018

[Hoffman 2018] J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. Efros and T. Darrell.
CyCADA: Cycle-consistent adversarial domain adaptation. ICML 2018

[Ganin 2015] Y. Ganin and V. Lempitsky.
Unsupervised domain adaptation by backpropagation. ICML 2015.

[Hoffman 2016] J. Hoffman, D. Wang, F. Yu, and T. Darrell.
FCNs in the wild: Pixel-level adversarial and constraint-based adaptation. arXiv 2016

[Tzeng 2017] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell.
Adversarial discriminative domain adaptation. CVPR 2017.

[Chen 2017] Y.-H. Chen, W.-Y. Chen, Y.-T. Chen, B.-C. Tsai, Y.-C. F. Wang, and M. Sun.
No more discrimination: Cross city adaptation of road scene segmenters. ICCV 2017

[Tsai 2018] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang and M. Chandraker.
Learning to adapt structured output space for semantic segmentation. CVPR 2018

[Lee 2019] K.-H. Lee, G. Ros, J. Li, and A. Gaidon.
SPIGAN: Privileged adversarial learning from simulation. ICLR 2019

[Vu 2019] T.-H. Vu, H. Jain, M. Bucher, M. Cord an P. Pérez.
AdvEnt: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation. CVPR 2019