Politecnico di Torino

1859

Adversarial Domain Adaptation for Real-time Semantic Segmentation

Giulia D’Ascenzi, Patrizio de Girolamo, Carlos Rosero

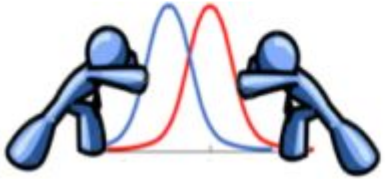
Advanced Machine Learning
2021-2022

Introduction

Semantic Segmentation

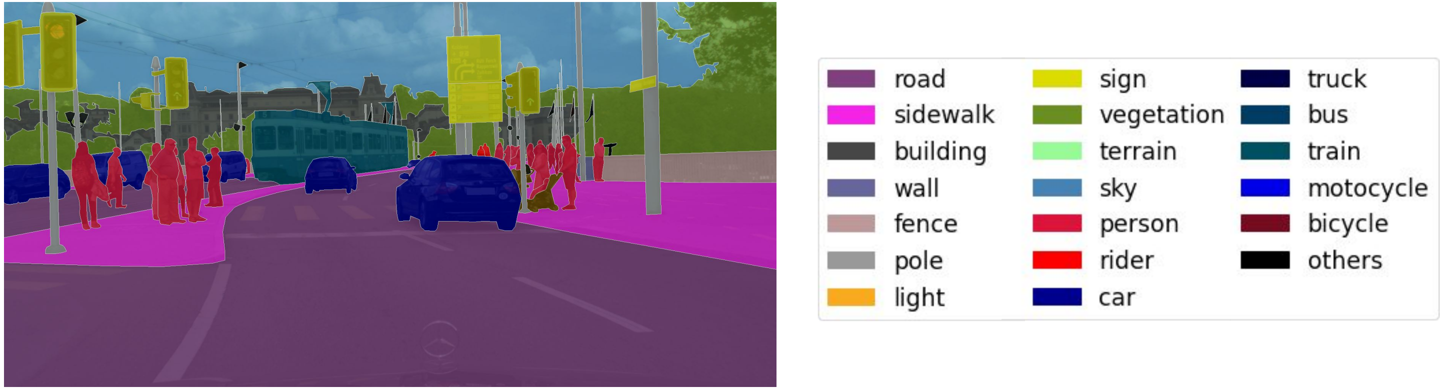
Real - Time applications

Domain Adaptation



Real - Time **Semantic Segmentation**

Goal: Assign to each pixel of an input image a category label



Possible Applications: Autonomous vehicles, virtual reality, computer - aided diagnosis, etc..

Real - Time Semantic Segmentation

Requested characteristics:

- ➔ Time - efficient inference speed
- ➔ Low memory and resources usage
- ➔ Good segmentation performances

⚠ Not something we can get with the standard semantic - segmentation models → Lots of parameters and Floating Point Operations (FLOP)

Real - Time Semantic Segmentation

Additional drawback: Semantic Segmentation models need to be trained on a large amount of densely labeled images.

It requires a large amount of human effort (time and money).

[1]: M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

Exploiting synthetic dataset

Idea: Train the networks on large photo - realistic synthetic datasets with computer - generated annotations.

GTA5 Dataset [2]

[2]: Richter S.R., Vineet V., Roth S., Koltun V. (2016) Playing for Data: Ground Truth from Computer Games. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9906. Springer, Cham.

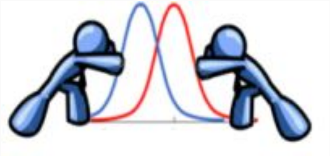
Domain Shift

GTA5
(Source)

Cityscapes
(Target)

Domain Shift

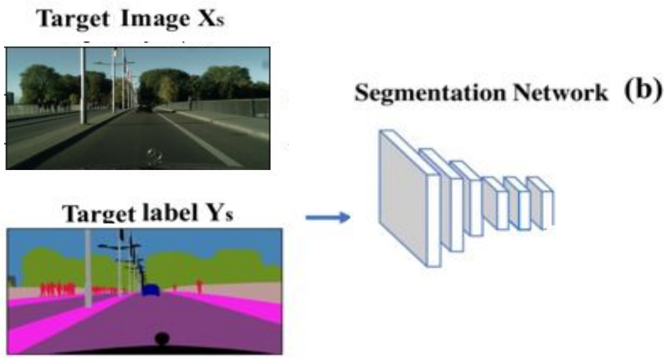
Domain Adaptation



GTA5
(Source)

Cityscapes
(Target)

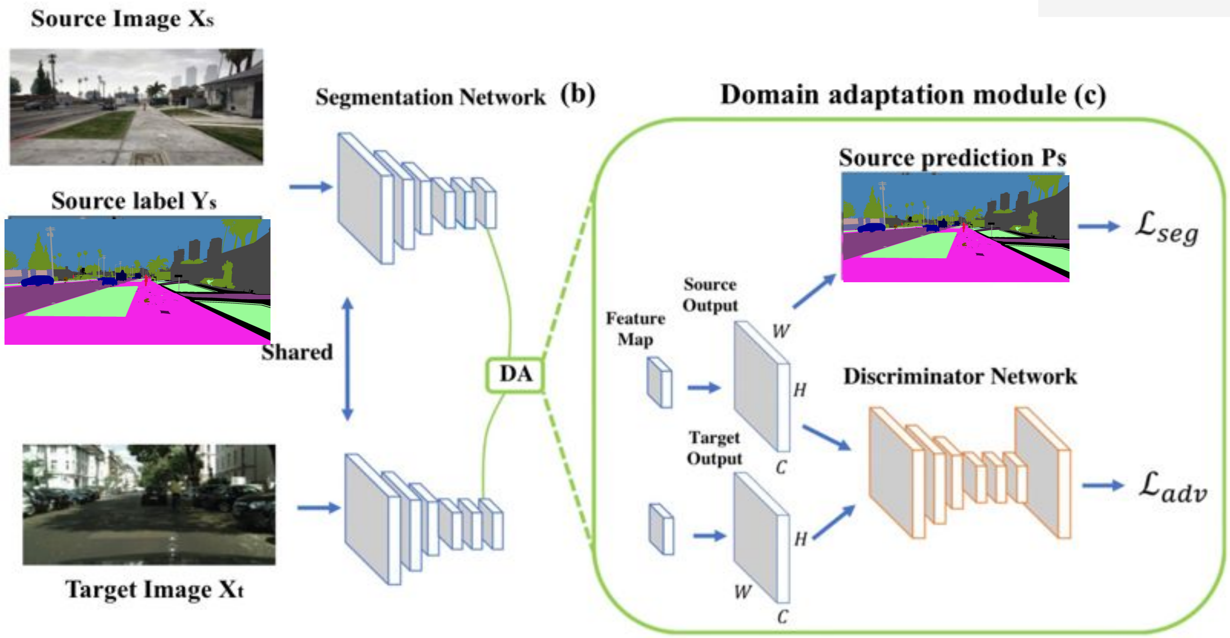
Overview Complete Architecture



1. Real- Time semantic Segmentation Network: BiSeNet [3]. Upper bound with target only.

[3] Yu C., Wang J., Peng C., Gao C., Yu G., Sang N. (2018) BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, vol 11217. Springer

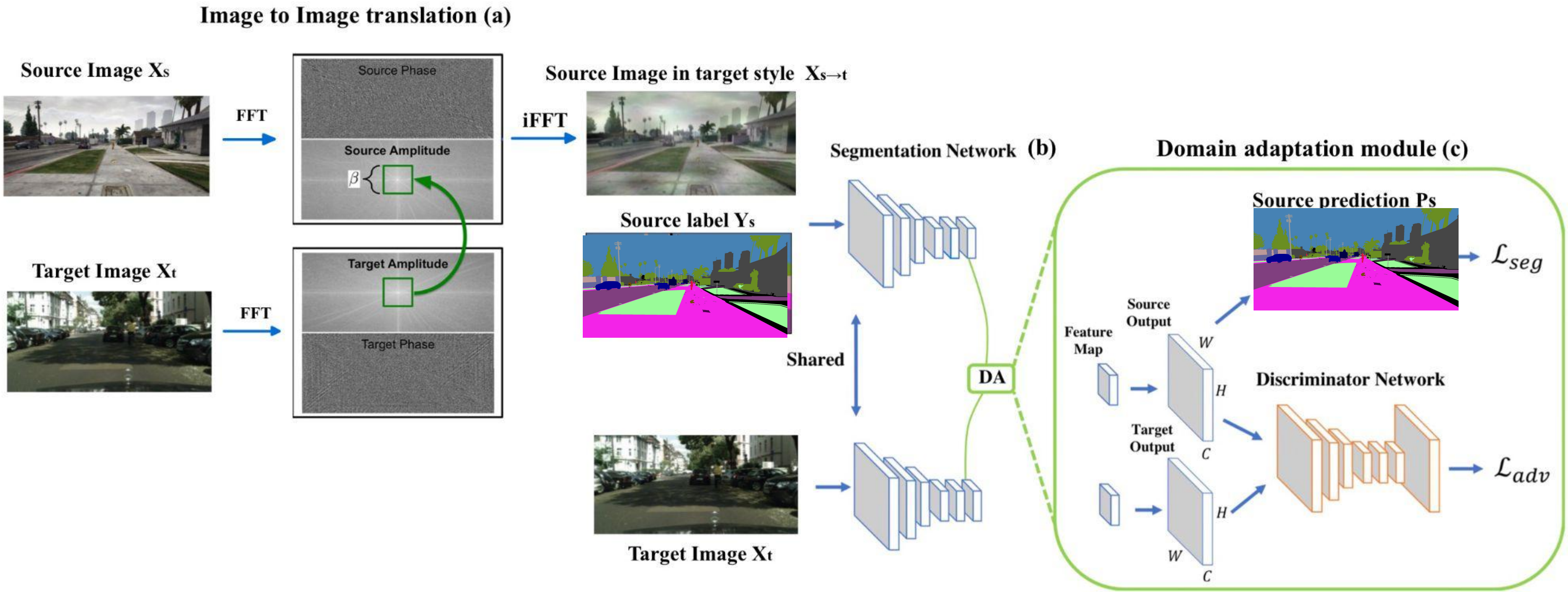
Overview Complete Architecture



2. Adversarial Domain Adaptation Module [4]

[5]Tsai, Yi-Hsuan, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang and Manmohan Chandraker. “Learning to Adapt Structured Output Space for Semantic Segmentation.” *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018): 7472-7481.

Overview Complete Architecture



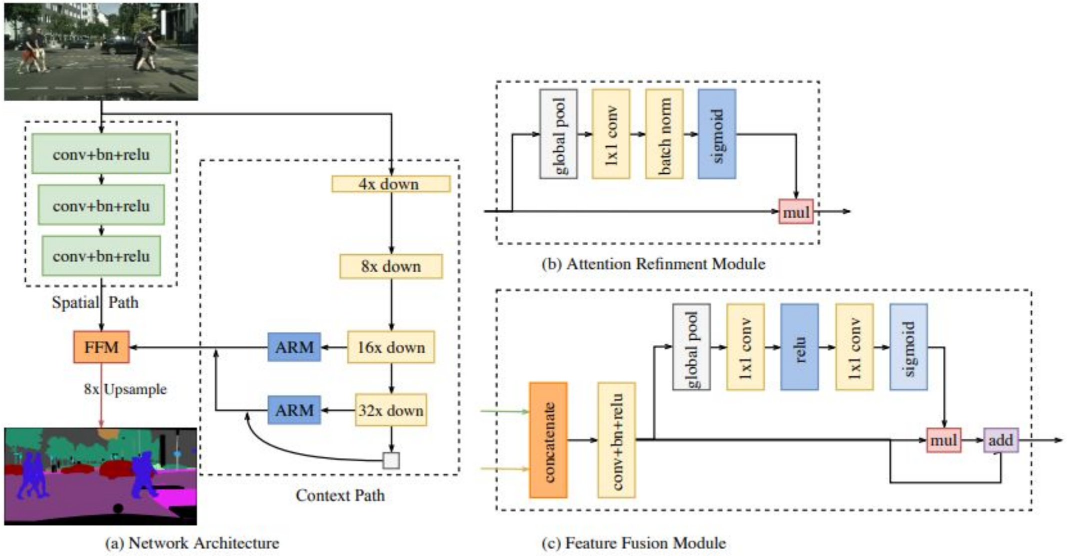
3. Increasing the performances: FDA [5]

[5]Yang, Yanchao and Stefano Soatto. "FDA: Fourier Domain Adaptation for Semantic Segmentation." *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020): 4084-4094.

Method

Semantic Segmentation

Network used: BiSeNet [3] (backbone Resnet-18 pretrained on ImageNet)



Capture semantics

Context path

Preserve spatial details

Spatial path

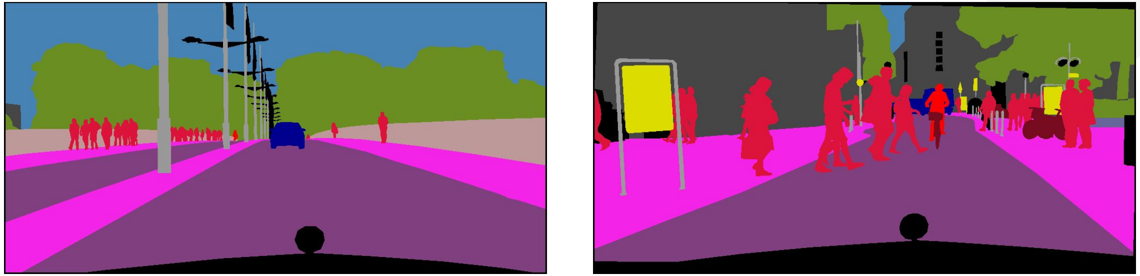
[3]: Yu C., Wang J., Peng C., Gao C., Yu G., Sang N. (2018) BiSeNet: Bilateral Segmentation Network for Real-Time Semantic Segmentation. In: Ferrari V., Hebert M., Sminchisescu C., Weiss Y. (eds) Computer Vision – ECCV 2018. ECCV 2018. Lecture Notes in Computer Science, vol 11217. Springer

(Without) Domain adaptation

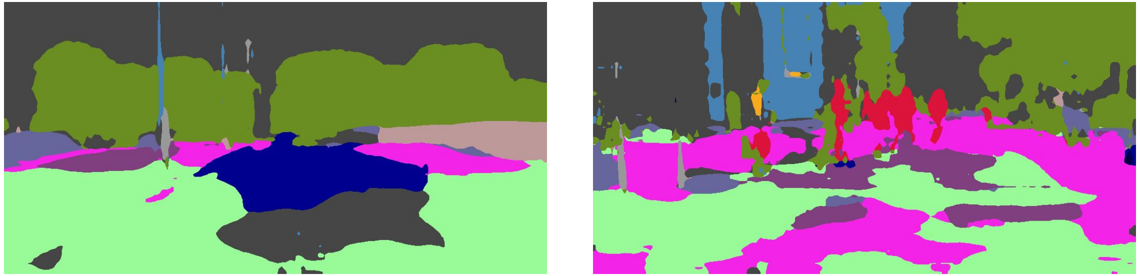
Goal: Transfer the knowledge obtained from the source finely annotated dataset to the target and unlabelled dataset

First try: Training BiSeNet using only GTA and then testing it on Cityscapes

Ground Truth:

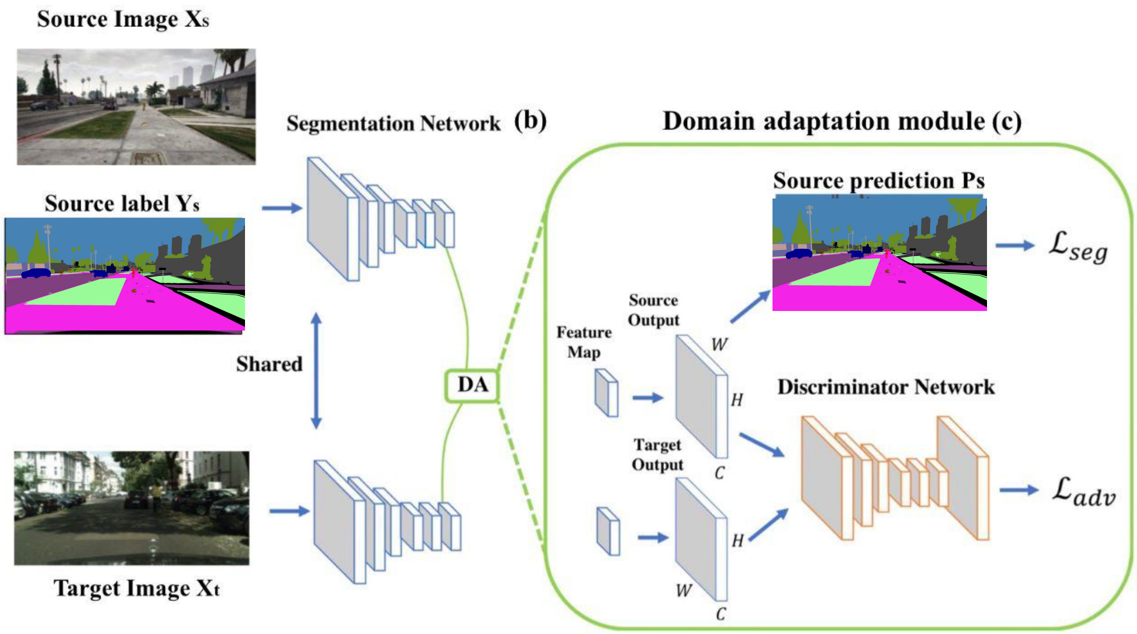


Output without DA:



Domain adaptation

Better Idea: Reducing the domain shift between source and target dataset using Unsupervised Adversarial Domain Adaptation



Unsupervised adversarial domain adaptation

Method: Min-Max game between the **generator** and the **discriminator**

Generator training:

$$\boxed{\mathcal{L}\left(X_s, X_t\right) = \mathcal{L}_{seg}\left(X_s\right) + \lambda_{adv}\mathcal{L}_{adv}\left(X_t\right)}$$

Where:

$$\mathcal{L}_{seg}\left(I_s\right) = -\sum_{h,w}\sum_{c\in C} Y_s^{(h,w,c)} \log\left(P_s^{(h,w,c)}\right)$$

$$\mathcal{L}_{adv}\left(I_t\right) = -\sum_{h,w} \log\left(\mathbf{D}\left(P_t\right)^{(h,w,1)}\right)$$

Discriminator Training:

$$\mathcal{L}_d(P) = -\sum_{h,w}(1-z)\log\left(\mathbf{D}(P)^{(h,w,0)}\right) + z\log\left(\mathbf{D}(P)^{(h,w,1)}\right)$$

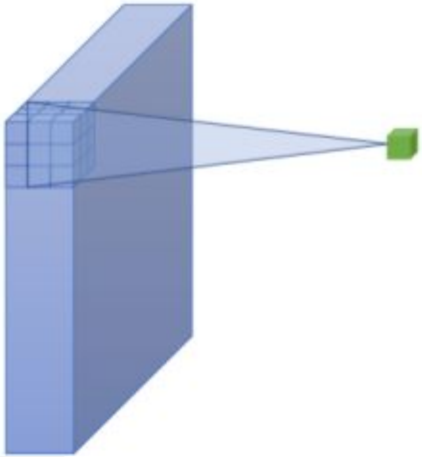
Discriminators

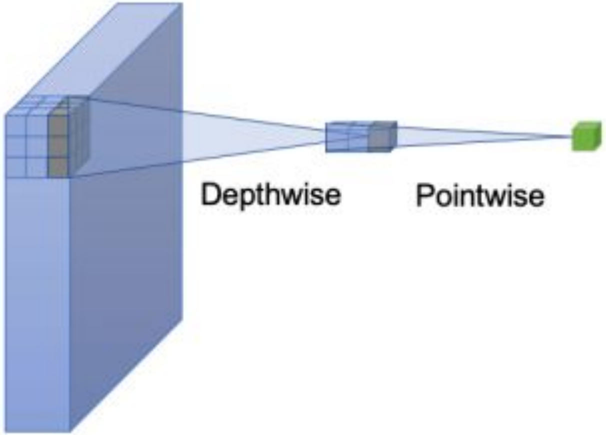
Fully Convolutional Discriminator vs **Lighter Discriminator**

Fully Convolutional Discriminator	Lighter Discriminator
2D convolutions	Depth-wise separable convolutions
Total Params: 2.7M Total FLOPS: 61.8G	Total Params: 191K Total FLOPS: 4.36G

Different types of convolution

2D convolutions vs **Depth-wise separable convolutions**





Increasing Performance: FDA

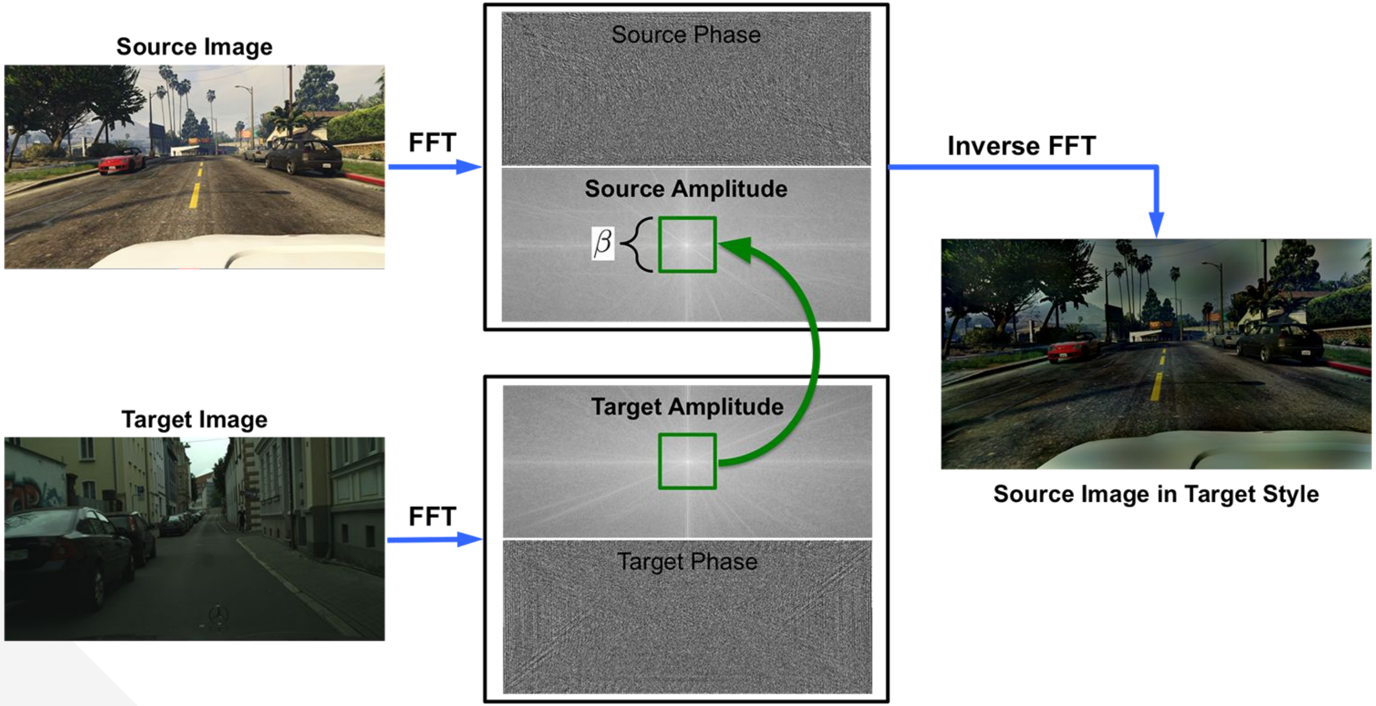
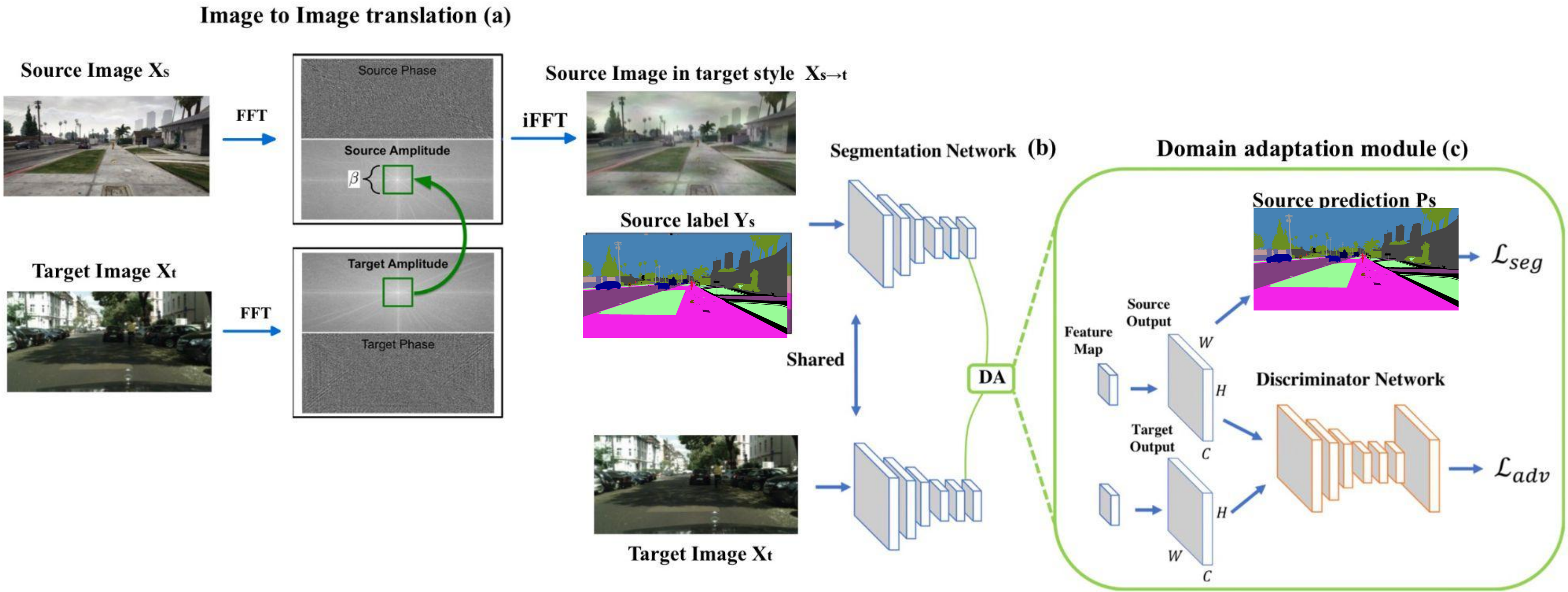


Figure Source: [22]

FDA β Ablation Study

Overview Complete Architecture



Results

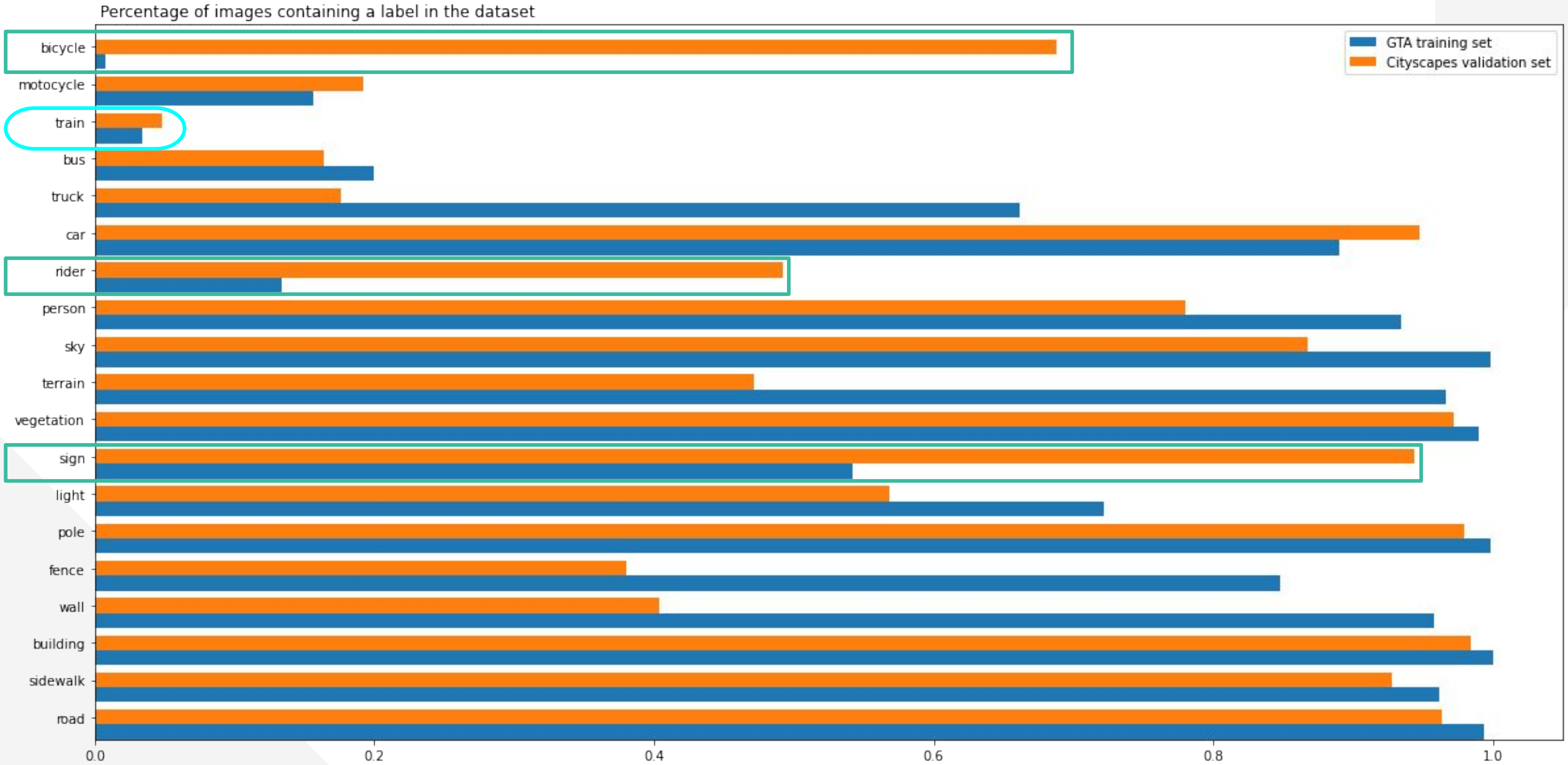
Results - Upper Bound / Augmentation

Augmentation	Accuracy(%)	mIoU(%)
None	79.3	47.3
Blur	79.1	47.0
Horizontal Flipping	79.8	49.7
Horizontal Flipping and Blur	79.7	**49.9**

Results

Method	mIoU %
Target only	**49.9**
No domain adaptation	13.4
Fully Conv	24.2
LW Conv	24.4
LW Conv, FDA ($\beta = 0.01$)	27.0
LW Conv , FDA ($\beta = 0.05$)	26.6
LW Conv, FDA ($\beta = 0.10$)	**27.4**
LW Conv , FDA ($\beta = 0.10$, no blur)	26.9

Results - Analysis



Results - Analysis

Method	Sign	Rider	Train	Bicycle	mIoU %
Target only	**45.9**	**30.2**	**25.9**	**53.3**	**49.9**
No domain adaptation	0.0	0.0	0.0	0.0	13.4
Fully Conv	0.1	0.1	0.0	0.0	24.2
LW Conv	0.2	1.0	0.0	0.0	24.4
LW Conv, FDA ($\beta = 0.01$)	1.4	1.4	0.0	0.0	27.0
LW Conv , FDA ($\beta = 0.05$)	0.6	8.2	0.0	0.0	26.6
LW Conv, FDA ($\beta = 0.10$)	**1.4**	**3.5**	**0.0**	**0.0**	**27.4**
LW Conv , FDA ($\beta = 0.10$, no blur)	1.3	4.5	0.0	0.0	26.9

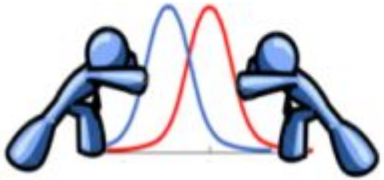
Conclusion

Semantic Segmentation

Real-Time Applications

Fourier Transform

Narrower Domain Gap

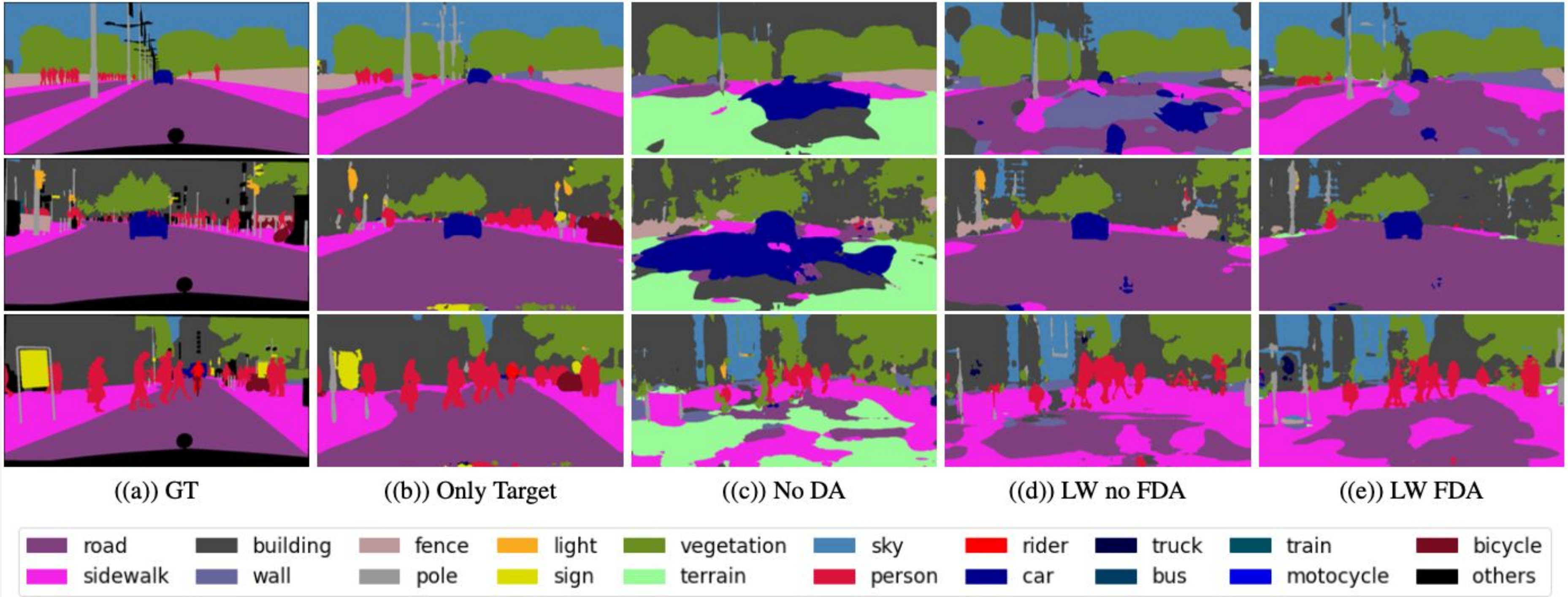


Thank You

Any Questions?

Appendix

Visual Comparison of Results



((a)) GT ((b)) Only Target ((c)) No DA ((d)) LW no FDA ((e)) LW FDA