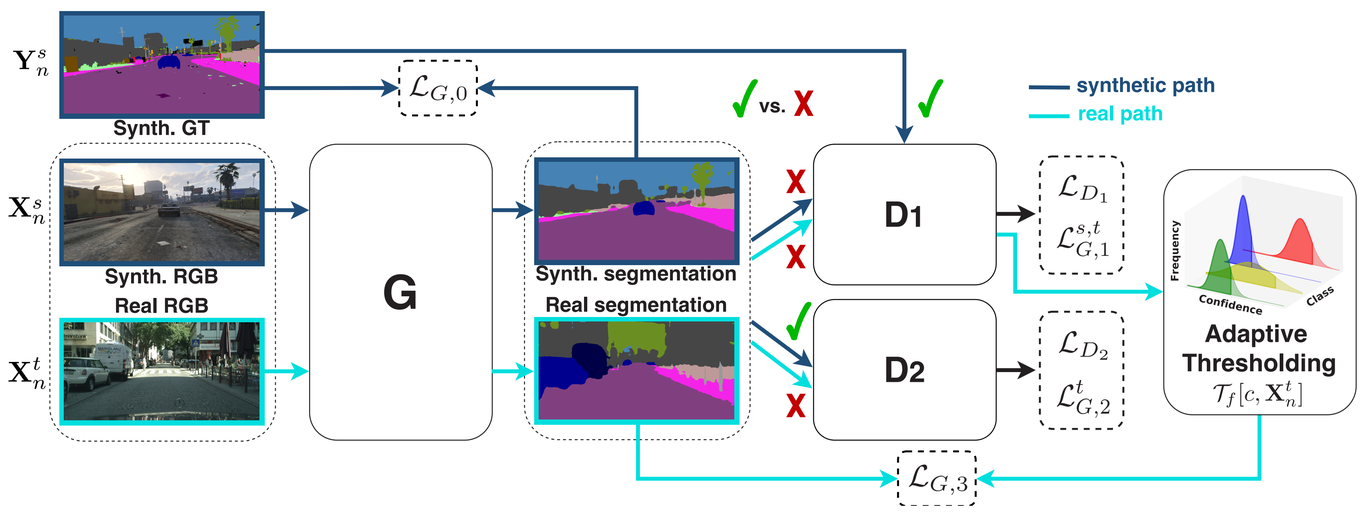
UNSUPERVISED DOMAIN ADAPTATION WITH MULTIPLE DOMAIN DISCRIMINATORS AND ADAPTIVE SELF-TRAINING

Teo Spadotto, Marco Toldo, Umberto Michieli and Pietro Zanuttigh
{toldomarco, michieli, zanuttigh}@dei.unipd.it
lttm.dei.unipd.it/paper_data/semanticDA/

Abstract

In this paper, we consider the semantic segmentation of urban scenes and propose an approach to adapt a deep neural network trained on synthetic data to real scenes. We introduce a novel UDA framework where a standard supervised loss on labeled synthetic data is supported by an adversarial module and a self-training strategy aiming at aligning the two domain distributions. The adversarial module is driven by a couple of fully convolutional discriminators dealing with different domains: the first discriminates between ground truth and generated maps, while the second between segmentation maps coming from synthetic or real world data. The self-training module exploits the confidence estimated by the discriminators on unlabeled data to select the regions used to reinforce the learning process. The confidence is further thresholded with an adaptive mechanism based on the per-class overall confidence. Experimental results prove the efficacy of the proposed strategy in adapting a segmentation network from synthetic datasets to real world ones.

Architecture of the Proposed Approach



Double Adversarial Adaptation

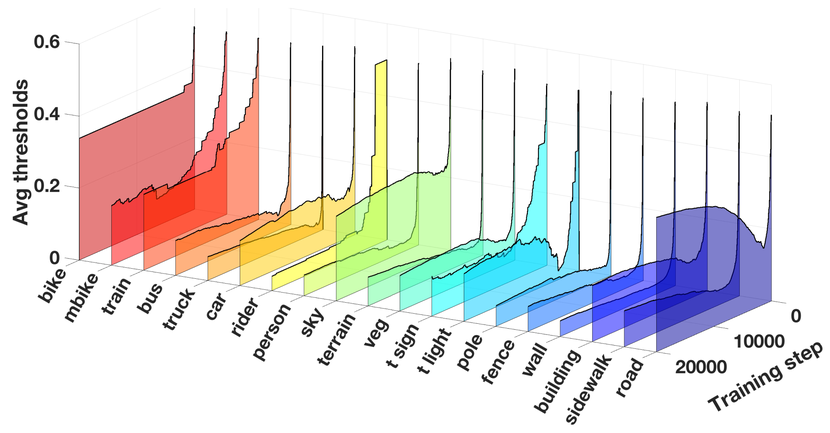
- D_1: discriminate between ground-truth segmentation maps and network predictions on both source and target input data, **indirect domain alignment**
- D_2: discriminate between source and target prediction maps, **direct domain alignment**

Self-Training

$$\mathcal{L}_{G,3} = -\sum_{p \in \mathbf{X}_n^t} \sum_{c \in \mathcal{C}} \mathcal{M}_f^{(p)} \cdot W_c^s \cdot \hat{\mathbf{Y}}_n^{(p)}[c] \cdot \log\left(G(\mathbf{X}_n^t)^{(p)}[c]\right)$$

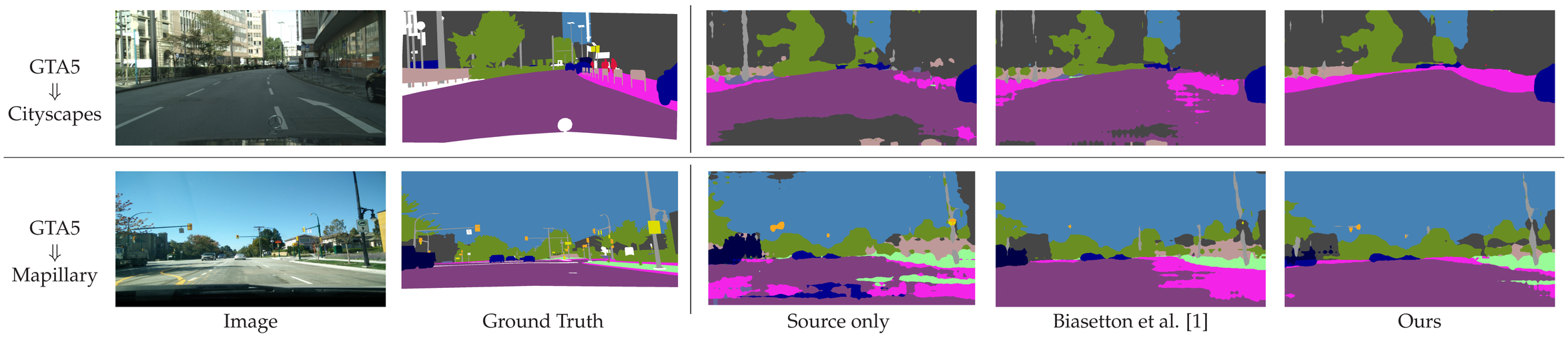
- Implicit adaptation from source
- Confidence selection from discriminator's output

Class-wise Confidence Threshold



Class and *step* adaptive thresholding dynamically updated

Qualitative Results: Image Segmentation



Quantitative Results

mIoU **GTA ⇒ CS**	road	sidewalk	building	wall	fence	pole	t light	t sign	veg	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mean
Source only	49.3	24.4	56.4	6.5	**19.6**	25.6	**23.6**	**10.1**	82.7	28.5	69.9	**55.5**	4.9	80.9	18.0	33.0	1.2	**15.1**	0.1	31.9
Ours	77.7	**35.9**	**67.2**	18.9	12.1	**26.2**	15.9	5.9	**83.7**	**33.3**	**72.7**	53.9	4.2	**82.6**	21.5	41.1	0.1	13.9	0.0	**35.1**
Biasetton et al. [1]	54.9	23.8	50.9	16.2	11.2	20.0	3.2	0.0	79.7	31.6	64.9	52.5	**7.9**	79.5	27.2	**41.8**	0.5	10.7	**1.3**	30.4
Michieli et al. [2]	**81.0**	19.6	65.8	**20.7**	2.9	20.9	6.6	0.2	82.4	33.0	68.2	54.9	6.2	80.3	**28.1**	41.6	**2.4**	8.5	0.0	33.3

mIoU **GTA ⇒ MP**	road	sidewalk	building	wall	fence	pole	t light	t sign	veg	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mean
Source only	69.8	31.8	58.8	14.6	22.3	28.3	**31.8**	**28.8**	70.0	24.4	72.4	**60.4**	16.8	80.6	36.6	34.3	**10.2**	**26.2**	**0.2**	37.8
Ours	**80.0**	**43.3**	**75.4**	19.4	**29.7**	**29.6**	23.3	16.2	78.5	33.5	93.7	59.0	**20.3**	**82.2**	**44.5**	**43.4**	2.5	22.1	0.0	**41.9**
Biasetton et al. [1]	71.4	25.0	62.0	20.4	17.6	26.8	5.9	0.8	64.6	24.6	86.5	58.3	14.7	80.0	39.3	42.2	5.5	22.3	0.1	35.2
Michieli et al. [2]	79.9	28.0	73.4	**23.0**	29.5	20.9	1.1	0.0	**79.5**	**39.6**	**95.0**	57.6	9.0	80.6	41.5	40.1	7.4	24.8	0.1	38.5

[1] M. Biasetton *et al.*, "Unsupervised Domain Adaptation for Semantic Segmentation of Urban Scenes," *CVPRW*, 2019.
[2] U. Michieli *et al.*, "Adversarial learning and self-teaching techniques for domain adaptation in semantic segmentation," in *TIV*, 2020.

Ablation Study

$\mathcal{L}_{G,0}$	$\mathcal{L}_{G,1}^s$	$\mathcal{L}_{G,1}^t$	$\mathcal{L}_{G,2}^t$	$\mathcal{L}_{G,3}$	$\mathcal{T}_f$	mIoU
✓						37.8
✓		✓	✓	✓	✓	39.9
✓	✓		✓	✓	✓	40.3
✓	✓	✓		✓	✓	40.7
✓	✓	✓	✓			41.1
✓	✓	✓	✓	✓		40.6
✓	✓	✓	✓	✓	fix 0.2	40.9
✓	✓	✓	✓	✓	✓	**41.9**

- All the components bring a significant contribution
- Self-training effective with confidence thresholds variable over both classes and training time