Supplementary Material
Beyond deterministic translation for unsupervised domain adaptation

Eleni Chiou
University College London
eleni.chiou.17@ucl.ac.uk

Iasonas Kokkinos
University College London
i.kokkinos@cs.ucl.ac.uk

Eleftheria Panagiotaki
University College London
panagio@cs.ucl.ac.uk

1. Quantitative Results

As we mentioned in the main paper, for the source domain network we observed experimentally that we obtained better results by adding an entropy-based regularization $\mathcal{L}_{Ent}$, to the output of source-domain network F_s when it is driven by translated target images. In Table 1 we report results obtained with and without the entropy-based regularization.

We further compare our approach with additional state-of-the-art methods [13, 9, 1, 4, 3, 15, 8, 12, 10, 7, 6, 11, 14]. We summarize the results for the **GTA-to-Cityscapes** benchmark in Table 2. All methods use DeepLabV2 [2] with ResNet101 [5] backbone.

Loss	mIoU
$\mathcal{L}_{CE}$	43.81
$\mathcal{L}_{CE} + \mathcal{L}_{Ent}$	**44.15**

Table 1: We obtain better results by adding a entropy-based regularization $\mathcal{L}_{Ent}$, to the output of source-domain network F_s when it is driven by translated target images.

2. Qualitative Results

Fig. 1 shows diverse translations of images from the SYNTHIA source dataset to the Cityscapes target dataset. We observe that our method generates sharp samples of high variability and noticeable diversity which results in substantially improved segmentation performance of the target-domain network.

Fig. 2 and Fig. 3 show diverse translations of images from the Cityscapes target dataset to the SYNTHIA and GTA source datasets respectively. We observe that our method generates diverse samples which preserve the content of the original image. By averaging the predictions of the source-domain network on multiple translated versions of the same target image allows us to estimate robust pseudo-labels for the the target dataset.

Fig. 4 and Fig. 5 show two examples of target-domain images translated to multiple source-domain images and labeled by the source-domain networks F_s. The label maps obtained by averaging the results obtained by different translations cancel out the fluctuation of the predictions around their ground-truth value and provide better pseudo-lables for the target dataset.

References

[1] Wei-Lun Chang, Hui-Po Wang, Wen-Hsiao Peng, and Wei-Chen Chiu. All about structure: Adapting structural information across domains for boosting semantic segmentation. In *CVPR*, June 2019. 1, 3

[2] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *TPAMI*, 2017. 1

[3] M. Chen, H. Xue, and D. Cai. Domain adaptation for semantic segmentation with maximum squares loss. In *ICCV*, 2019. 1, 3

[4] Liang Du, Jingang Tan, Hongye Yang, Jianfeng Feng, Xiangyang Xue, Qibao Zheng, Xiaoqing Ye, and Xiaolin Zhang. Ssf-dan: Separated semantic feature based domain adaptation network for semantic segmentation. In *ICCV*, 2019. 1, 3

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 1

[6] Jiaxing Huang, Shijian Lu, Dayan Guan, and Xiaobing Zhang. Contextual-relation consistent domain adaptation for semantic segmentation. In *ECCV*, 2020. 1, 3

[7] Guangrui Li, Guoliang Kang, Wu Liu, Yunchao Wei, and Yi Yang. Content-consistent matching for domain adaptive semantic segmentation. In *ECCV*, 2020. 1, 3

[8] Qing Lian, Fengmao Lv, Lixin Duan, and Boqing Gong. Constructing self-motivated pyramid curriculums for cross-domain semantic segmentation: A non-adversarial approach. In *ICCV*, 2019. 1, 3

[9] Y. Luo, L. Zheng, T. Guan, J. Yu, and Y. Yang. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *CVPR*, 2019. 1, 3

[10] Fengmao Lv, Tao Liang, Xiang Chen, and Guosheng Lin. Cross-domain semantic segmentation via domain-invariant interactive relation transfer. In *CVPR*, June 2020. 1, 3

[11] Haoran Wang, Tong Shen, Wei Zhang, Lingyu Duan, and Tao Mei. Classes matter: A fine-grained adversarial approach to cross-domain semantic segmentation. In *ECCV*, 2020. 1, 3

[12] Zhonghao Wang, Mo Yu, Yunchao Wei, Rogerio Feris, Jinjun Xiong, Wen-mei Hwu, Thomas S. Huang, and Honghui Shi. Differential treatment for stuff and things: A simple unsupervised domain adaptation method for semantic segmentation. In *CVPR*, 2020. 1, 3

[13] Zuxuan Wu, Xintong Han, Yen-Liang Lin, Mustafa Gokhan Uzunbas, Tom Goldstein, Ser Nam Lim, and Larry S Davis. Dcan: Dual channel-wise alignment networks for unsupervised scene adaptation. In *ECCV*, 2018. 1, 3

[14] Jinyu Yang, Weizhi An, Sheng Wang, Xinliang Zhu, Chaochao Yan, and Junzhou Huang. Label-driven reconstruction for domain adaptation in semantic segmentation. In *ECCV*, 2020. 1, 3

[15] Yang Zou, Zhiding Yu, B. V. K. Vijaya Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *ECCV*, 2018. 1, 3

Method	*road*	*sidewalk*	*building*	*wall*	*fence*	*pole*	*light*	*sign*	*vegetation*	*terrain*	*sky*	*person*	*rider*	*car*	*truck*	*bus*	*train*	*motocycle*	*bicycle*	mIoU
ResNet backbone																				
DCAN [13]	85.0	30.8	81.3	25.8	21.2	22.2	25.4	26.6	83.4	36.7	76.2	58.9	24.9	80.7	29.5	42.9	2.5	26.9	11.6	41.7
CLAN [9]	88.5	35.4	79.5	26.3	24.3	28.5	32.5	18.3	81.2	40.0	76.5	58.1	25.8	82.6	30.3	34.4	3.4	21.6	21.5	42.6
DISE [1]	91.5	47.5	82.5	31.3	25.6	33.0	33.7	25.8	82.7	28.8	82.7	62.4	30.8	85.2	27.7	34.5	6.4	25.2	24.4	45.4
Ssf-dan [4]	90.3	38.9	81.7	24.8	22.9	30.5	37.0	21.2	84.8	38.8	76.9	58.8	30.7	85.7	30.6	38.1	5.9	28.3	36.9	45.4
MSL [3]	89.4	43.0	82.1	30.5	21.3	30.3	34.7	24.0	85.3	39.4	78.2	63.0	22.9	84.6	36.4	43.0	5.5	34.7	33.5	46.4
CBST [15]	89.6	58.9	78.5	33.0	22.3	41.4	48.2	39.2	83.6	24.3	65.4	49.3	20.2	83.3	39.0	48.6	12.5	20.3	35.3	47.0
PyCDA [8]	90.5	36.3	84.4	32.4	28.7	34.6	36.4	31.5	86.8	37.9	78.5	62.3	21.5	85.6	27.9	34.8	18.0	22.9	49.3	47.4
Wang et al.[12]	90.6	44.7	84.8	34.3	28.7	31.6	35.0	37.6	84.7	43.3	85.3	57.0	31.5	83.8	42.6	48.5	1.9	30.4	39.0	49.2
PIT [10]	87.5	43.4	78.8	31.2	30.2	36.3	39.9	42.0	79.2	37.1	79.3	65.4	37.5	83.2	46.0	45.6	25.7	23.5	49.9	50.6
CCM [7]	93.5	57.6	84.6	39.3	24.1	25.2	35.0	17.3	85.0	40.6	86.5	58.7	28.7	85.8	49.0	56.4	5.4	31.9	43.2	49.9
CrCDA [6]	92.4	55.3	82.3	31.2	29.1	32.5	33.2	35.6	83.5	34.8	84.2	58.9	32.2	84.7	40.6	46.1	2.1	31.1	32.7	48.6
FADA-MST [11]	91.0	50.6	86.0	43.4	29.8	36.8	43.4	25.0	86.8	38.3	87.4	64.0	38.0	85.2	31.6	46.1	6.5	25.4	37.1	50.1
Yang et al. [14]	90.8	41.4	84.7	35.1	27.5	31.2	38.0	32.8	85.6	42.1	84.9	59.6	34.4	85.0	42.8	52.7	3.4	30.9	38.1	49.5
Ours	91.3	42.1	85.6	43.6	31.8	32.0	38.2	35.4	88.5	46.4	84.5	61.7	31.4	86.5	37.2	51.3	4.2	38.5	39.0	**51.0**

Table 2: Quantitative Comparison on GTA5→Cityscapes.

Figure 1: Diverse translations of images from the SYNTHIA source dataset to the Cityscapes target dataset: we observe that even though the content and pixel semantics stay intact, we generate diverse variants of the same scene, effectively capturing more faithfully the data distribution in the target domain.

Figure 2: Diverse translations of images from the Cityscapes target dataset to the SYNTHIA source dataset: we observe that even though the content and pixel semantics stay intact, we generate diverse variants of the same scene, effectively capturing more faithfully the data distribution in the source domain.

Figure 3: Diverse translations of images from the Cityscapes target dataset to the GTA source dataset: we observe that even though the content and pixel semantics stay intact, we generate diverse variants of the same scene, effectively capturing more faithfully the data distribution in the source domain.

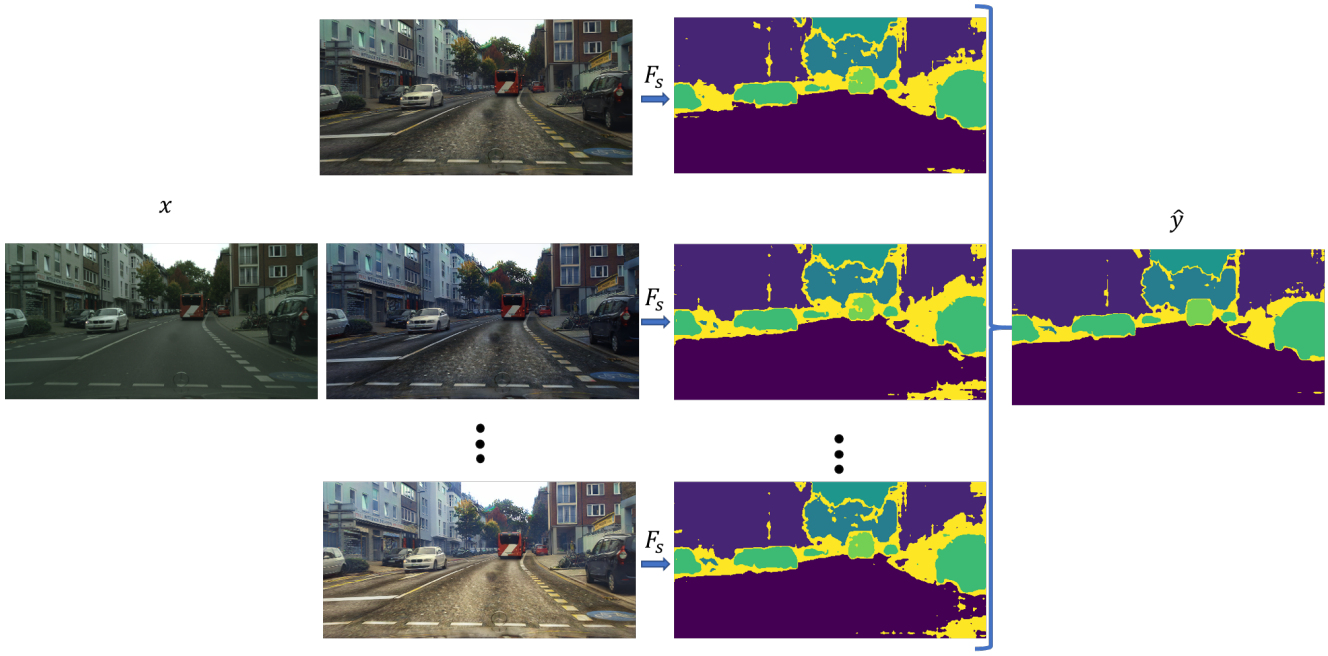

Figure 4: The target image (left) results in multiple target-domain translations (middle) which are processed by the source-domain network, F_s and averaged to produce pseudo-labels for the target image; the latter are used to supervise the target-domain network F_t through a cross-entropy loss.

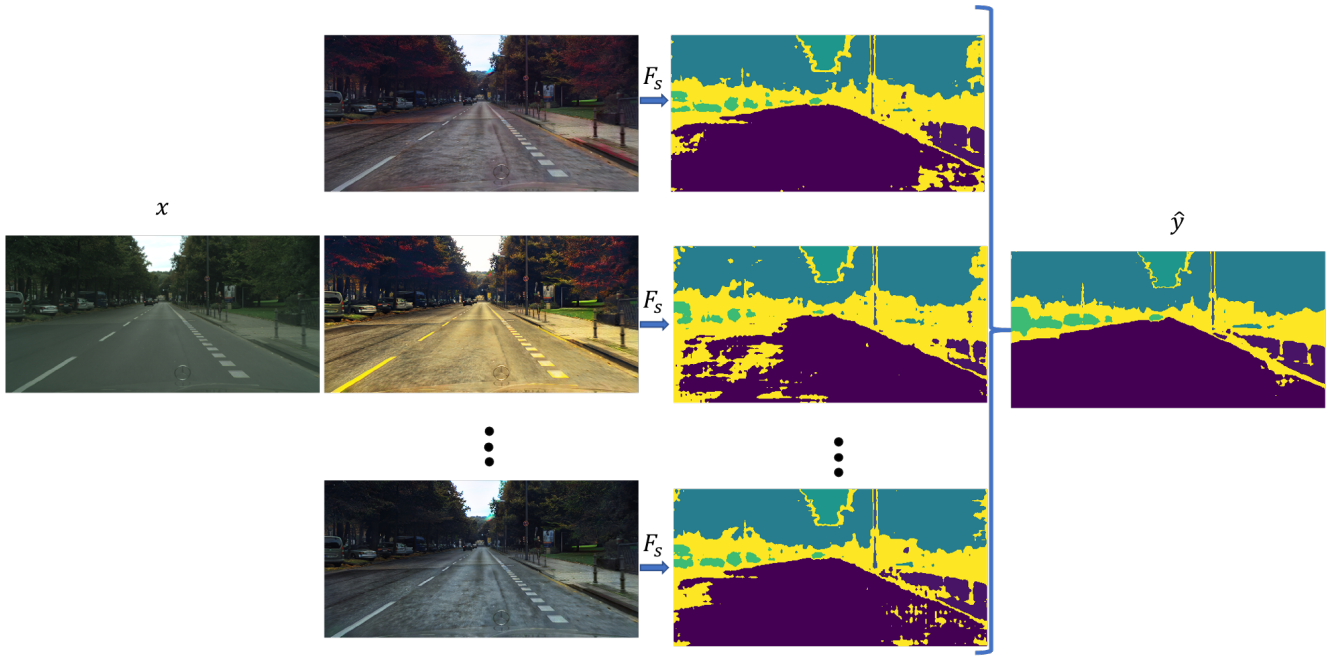

Figure 5: The target image (left) results in multiple target-domain translations (middle) which are processed by the source-domain network, F_s and averaged to produce pseudo-labels for the target image; the latter are used to supervise the target-domain network F_t through a cross-entropy loss.