Supplementary of :
Domain Bridge for Unpaired Image-to-Image Translation and Unsupervised Domain Adaptation

Fabio Pizzati[1,2], Raoul de Charette[2], Michela Zaccaria[3], Pietro Cerri[1]
[1]VisLab, [2]Inria, [3]University of Parma
{fabio.pizzati, raoul.de-charette}@inria.fr
michela.zaccaria@studenti.unipr.it
pcerri@ambarella.com

We report additional results on the clear→rain task of our MUNIT-bridged image-to-image translation (*i2i*, cf. article Sec. 3.1) and our unsupervised domain adaptation (*UDA*, cf. article Sec. 3.2). All results are obtained with the same training procedure and hyperparameters described in the article.

1. Bridged image-to-image (i2i) translation

In Figs. 1 and 2, more qualitative i2i translation outputs are compared to the recent CycleGAN [5] and MUNIT [2]. As for the results in the main paper, random images of the Cityscapes [1] validation set are processed with the i2i GANs.

It is noticeable that the results are consistent in variability and quality. Our methods (last 3 rows) is the only one that generates realistic rainy scenes, adding even accurate reflections. This is evident in the first columns of Fig. 1, where parked cars are clearly reflected on the wet ground. This shows the effectiveness of our domain bridge since the scene geometry is never explicitly input. Meanwhile, CycleGAN and MUNIT perform drastically worse (see explanation in the article).

Our method still tends to over saturate translated images with vivid color (e.g. Fig. 1, col 2, last row). However, drops on the windshield are correctly generated in all styles. In the last cols and rows of Fig. 2, some undesired artifacts are added by the GAN (green spots), probably due to road reflections in the training set that are encoded in some styles. Those are also appear in Fig. 1 (col 2, row 4), but in a less evident form. We assume that this happens only for a restricted styles set, since those are the only significant examples of this behavior in our sampling.

2. Unsupervised Domain Adaptation (UDA)

Figs. 3 and 4 show additional UDA segmentation results on BDD-rainy validation set (cf. article, Sec. 4.1.1). In addition to the complete *Ours + OMS + WPL* outputs using the Weighted Pseudo Labels (WPL, cf. article Sec. 3.2), we further include qualitative outputs from *Ours* approach, i.e. training on our domain-bridged i2i translations, and *Ours + OMS* that is with Online Multimodal Style-sampling (cf. article Sec. 3.2). Along with our results we report results from the very recent BDL [3] and AdaptSegNet [4].

Again, qualitative results are in fair alignment with mIoU metrics reported in article Tab. 2. Even the basic *Ours* approach leads to significant smoother segmentation w.r.t. baseline, particularly in regions containing drops and reflections, e.g. Fig. 3 col 5. This is also visible w.r.t. to BDL or AdaptSegNet, as the images with evident water drops are also the ones where our method outperforms the others more significantly (e.g. Fig. 3 col 5). Once again, this shows the effectiveness of the domain bridge in UDA. Qualitative effects vary for Weighted Pseudo Label (cf. article Sec. 3.2). For *Ours + OMS + WPL*, slight qualitative improvements is visible in some frames (Fig. 3, last column), while a few are negatively impacted, as in Fig. 3, first column.

Figure 1: Results on qualitative comparison between state-of-the-art architectures and our approach for i2i in the clear→rain transformation on Cityscapes validation set [1]. Our method is the only one that adds typical traits of rain as water drops and reflections.

Figure 2: Results on qualitative comparison between state-of-the-art architectures and our approach for i2i in the clear→rain transformation on Cityscapes validation set [1]. Our method is the only one that adds typical traits of rain as water drops and reflections.

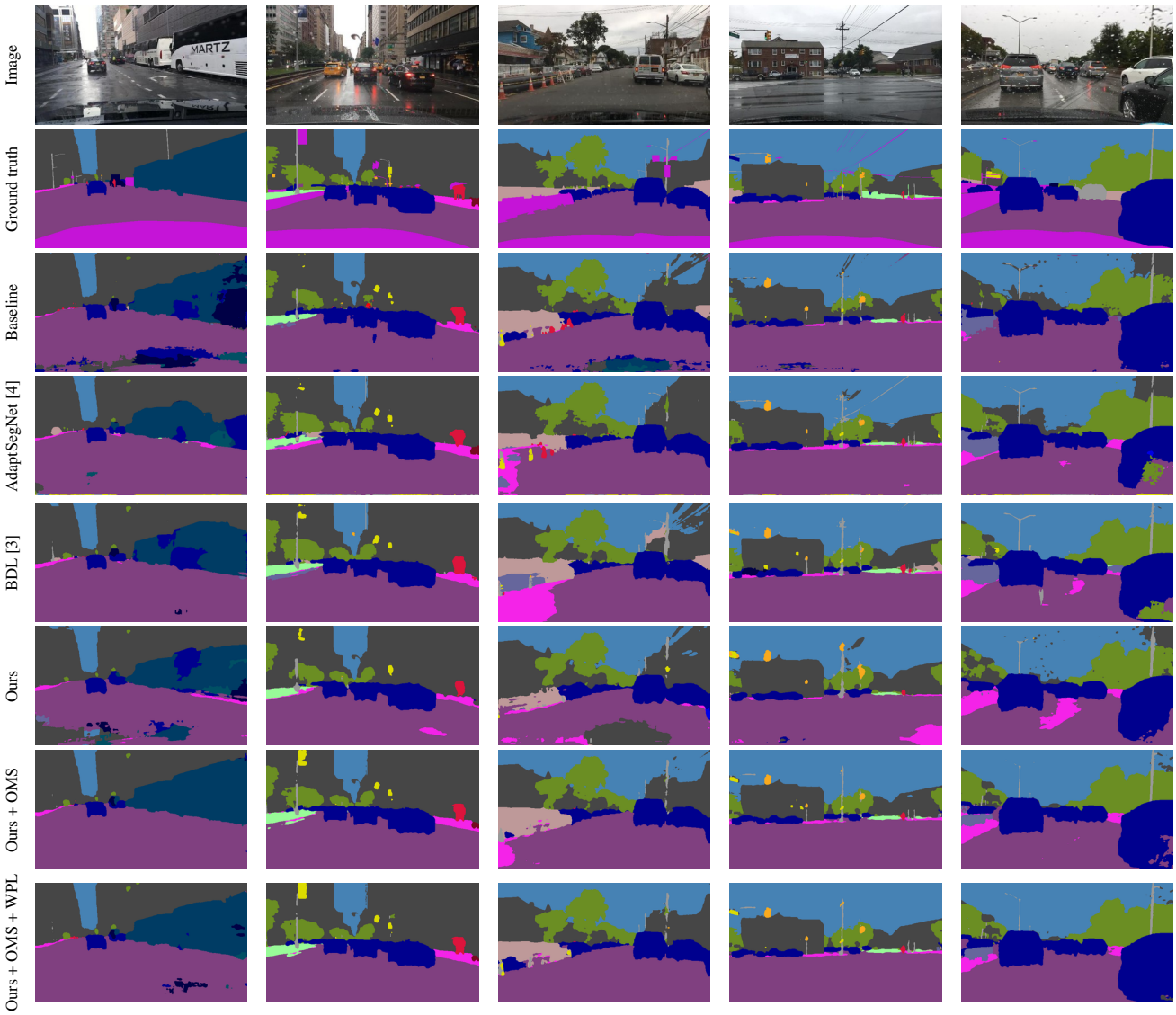

Figure 3: Comparison of our method with the state-of-the-art for semantic segmentation UDA on BDD-rainy validation set (cf. article, Sec. 4.1.1). We perform on par with the very recent BDL [3] and significantly outperform AdaptSegNet [4].

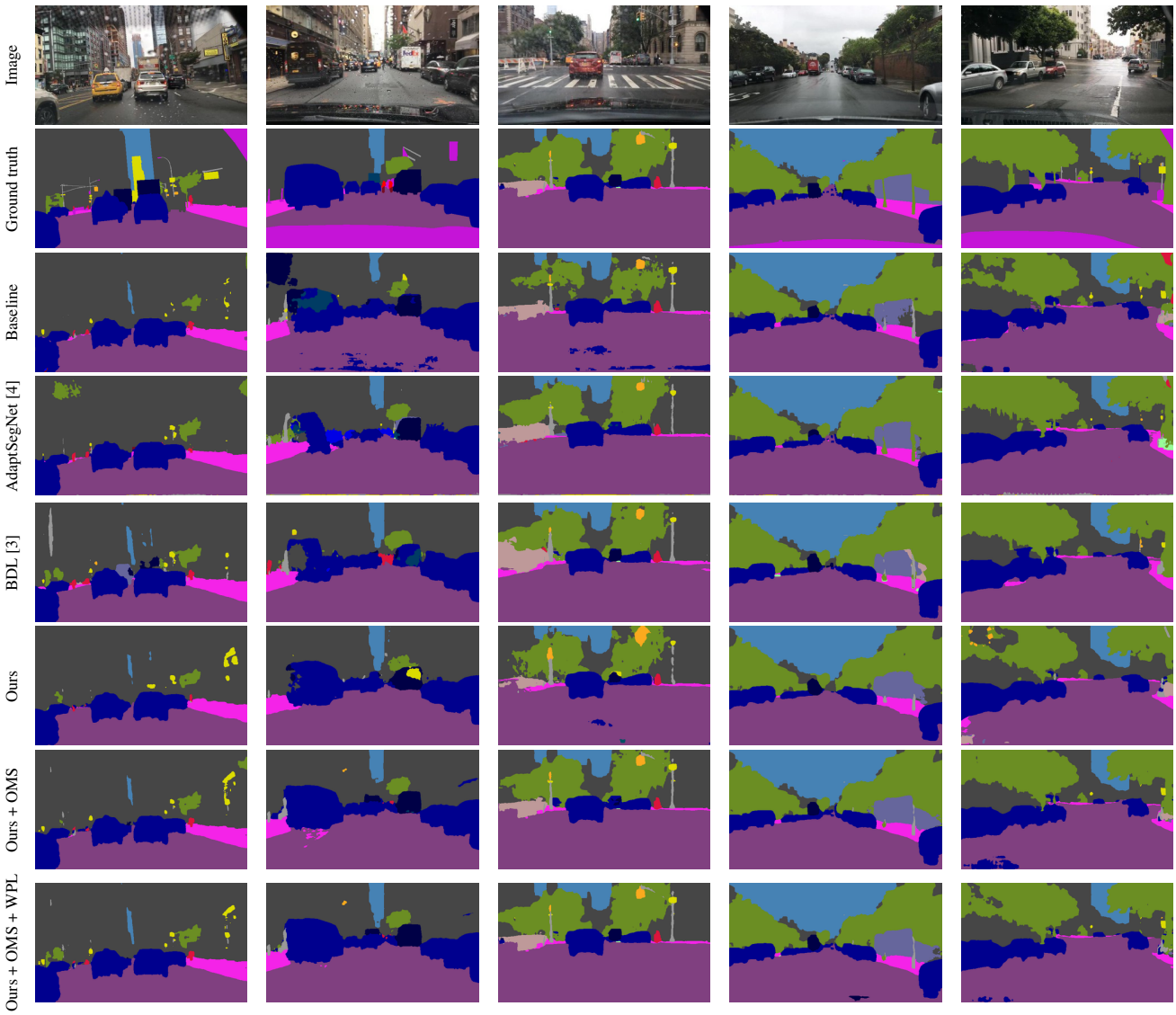

Figure 4: Comparison of our method with the state-of-the-art for semantic segmentation UDA on BDD-rainy validation set (cf. article, Sec. 4.1.1). We perform on par with the very recent BDL [3] and significantly outperform AdaptSegNet [4].

References

[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 *CVPR*, 2016.

[2] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz. Multimodal unsupervised image-to-image translation. In *ECCV*, 2018.

[3] Y. Li, L. Yuan, and N. Vasconcelos. Bidirectional learning for domain adaptation of semantic segmentation. In *CVPR*, 2019.

[4] 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. In *CVPR*, 2018.

[5] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017.