Spatial Class Distribution Shift in Unsupervised Domain Adaptation: Local Alignment Comes to Rescue
Supplementary Material

Safa Cicek[1] Ning Xu[2] Zhaowen Wang[2] Hailin Jin[2] Stefano Soatto[1]

[1] University of California, Los Angeles, {safacicek,soatto}@ucla.edu
[2] Adobe Research, {nxu,zhawang,hljin}@adobe.com

1 Additional Experiments on Berkeley Deep Driving

Table 1. All models are trained on the labeled source training data and unlabeled target training data and performances on the validation split of the target data are reported.

Cityscapes → Berkeley

Method	Road	SW	Build	Wall	Fence	Pole	TL	TS	Veg.	Terrain	Sky	PR	Rider	Car	Truck	Bus	Train	Motor	Bike	mIoU
Source-only	76.25	42.15	64.99	11.7	23.95	30.75	29.4	34.98	77.23	17.49	82.67	44.87	31.0	79.37	20.34	37.63	0.04	40.33	29.34	40.76
AGP-GI	87.77	45.89	71.79	14.63	28.27	31.73	35.59	38.5	78.29	27.24	84.37	46.7	31.89	82.36	28.59	32.48	0.0	32.9	26.17	43.43
Ours	90.79	50.06	72.57	13.08	29.53	37.45	36.46	43.01	82.81	33.33	85.93	55.23	39.26	85.28	28.12	40.37	0.0	41.8	33.35	**47.29**

SYNTHIA → Berkeley

Method	Road	SW	Build	Wall*	Fence*	Pole*	TL	TS	Veg.	Sky	PR	Rider	Car	Bus	Motor	Bike	mIoU	mIoU-13
Source-only	13.42	9.85	41.97	1.17	0.0	14.39	20.46	11.06	52.29	68.79	21.9	6.0	45.23	3.46	4.7	12.13	20.43	23.94
AGP-GI	3.82	7.67	45.34	1.52	0.06	17.99	16.99	8.19	58.21	64.6	16.38	5.91	61.86	6.06	9.64	22.86	21.69	25.19
Ours	29.07	11.15	57.82	1.56	0.0	27.44	30.67	14.22	65.64	78.48	32.85	16.88	67.85	20.34	24.73	33.69	**32.02**	**37.18**

GTA5 → Berkeley

Method	Road	SW	Build	Wall	Fence	Pole	TL	TS	Veg.	Terrain	Sky	PR	Rider	Car	Truck	Bus	Train	Motor	Bike	mIoU
Source-only	51.68	16.39	41.22	2.27	27.66	30.1	34.46	19.56	56.58	26.36	64.19	48.32	20.94	70.64	14.05	31.87	0.0	27.31	26.79	32.13
AGP-GI	81.16	14.42	66.86	9.13	28.33	32.22	36.63	26.89	68.01	24.1	83.27	49.74	28.48	77.67	19.09	17.91	0.0	34.05	33.87	38.52
Ours	81.73	25.97	69.45	12.69	32.12	34.83	40.02	29.75	70.43	30.58	83.62	56.23	28.57	80.64	27.9	52.3	0.0	36.08	32.04	**43.42**

Berkeley → Cityscapes

Method	Road	SW	Build	Wall	Fence	Pole	TL	TS	Veg.	Terrain	Sky	PR	Rider	Car	Truck	Bus	Train	Motor	Bike	mIoU
Source-only	93.93	58.07	84.52	27.29	34.38	36.07	36.13	45.43	85.95	46.81	85.05	59.2	32.36	88.27	50.99	52.98	0.03	26.74	47.37	52.19
AGP-GI	93.99	60.24	84.98	29.24	33.44	34.5	35.65	45.31	86.25	45.58	87.42	62.19	36.33	88.19	45.03	54.67	0.16	30.45	48.6	52.75
Ours	93.86	58.31	85.45	38.49	31.67	36.2	32.49	42.7	86.92	47.25	87.41	63.38	37.77	90.29	68.57	61.04	6.11	35.43	51.06	**55.5**

In this section, we evaluate the proposed local alignment method on the Berkeley Deep Driving dataset [1]. Berkeley Deep Driving dataset consists of drive-cam images with resolution 1280 × 720. Frames are collected at 30 FPS and each 10th frame is annotated. We used the extended version of this dataset by [1],

BDD100K and not the earlier version [2]. In this version, the number of samples are 7,000, 1,000 and 2,000 for training, validation and test splits respectively. We used the training split for training and the validation split for reporting the performance. We did not use the test split. Categories are compatible with Cityscapes and GTA5. The dataset covers diverse driving scenarios like different times of day (e.g. daytime, nighttime, dusk, dawn) and diverse weather conditions (e.g. sunny, rainy, snowy). The majority of data comes from New York, San Francisco, Berkeley unlike Cityscapes which covers various cities in Germany and neighboring countries making spatial layouts across datasets slightly different.

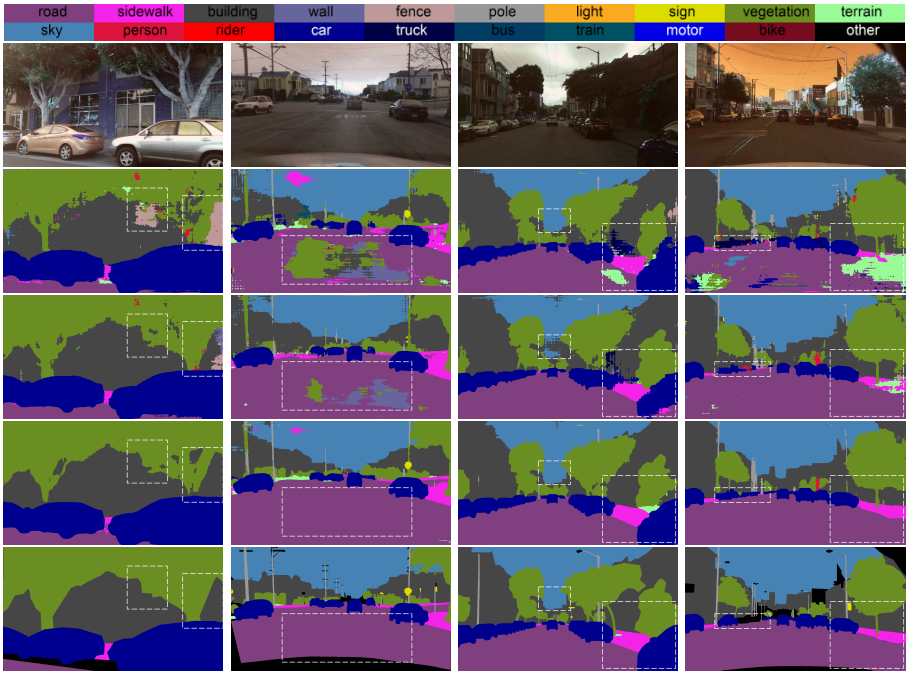

Fig. 1. Cityscapes → Berkeley. From top to bottom: (1) Image, (2) source-only prediction, (3) global alignment prediction, (4) our prediction and (5) ground truth segmentation. Best viewed in color.

In Table 1, we report the performances of the proposed method and the baselines on four different UDA settings namely: Cityscapes → Berkeley, SYNTHIA → Berkeley, GTA5 → Berkeley, and Berkeley → Cityscapes. The source-only model is only trained on the labeled source examples minimizing the cross-entropy loss. AGP-GI (Align Global Predictions of Global Images) refers to minimizing the same adversarial loss but on the global segmentation maps sim-

ilar to [3]. We follow the implementation details described in the main text for both the proposed method and the baseline methods.

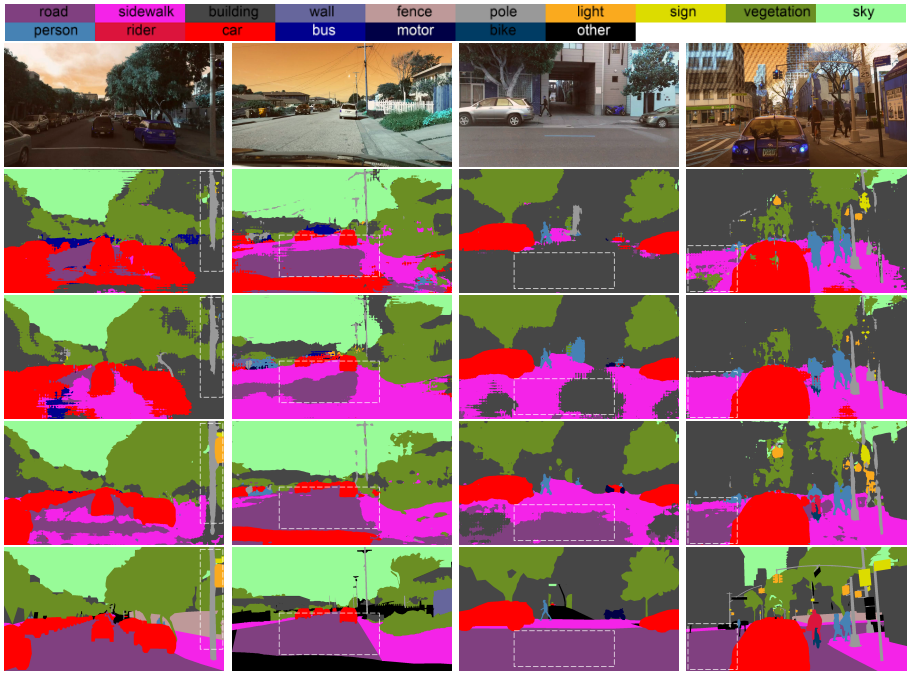

Fig. 2. SYNTHIA → Berkeley. From top to bottom: (1) Image, (2) source-only prediction, (3) global alignment prediction, (4) our prediction and (5) ground truth segmentation. Best viewed in color.

All the methods have higher scores on Berkeley → Cityscapes compared to Cityscapes → Berkeley. The transfer from Berkeley → Cityscapes is easier compared to Cityscapes → Berkeley as Berkeley covers more diverse scenes. Furthermore, mIoU scores reported in the main text for GTA5 → Cityscapes (46.98 %) and SYNTHIA → Cityscapes (51.99 %) are higher than GTA5 → Berkeley (43.43 %) and SYNTHIA → Berkeley (37.18 %). This was expected due to the larger diversity of the Berkeley dataset relative to Cityscapes. However, the proposed local alignment outperforms the baselines in all tasks. The proposed method surpasses the global-alignment baselines with 11.99 %, 4.9 %, 3.86 % and 2.75 % (mIoU) for SYNTHIA → Berkeley, GTA5 → Berkeley, Cityscapes → Berkeley and Berkeley → Cityscapes respectively. The proposed method especially shines on SYNTHIA → Berkeley where the spatial class distribution shift is the largest. All other datasets have dashcam views while SYNTHIA has random camera views.

[2,4,5] reported on a single one of these four tasks. But, possibly they run on an earlier version of the dataset, so to not have an unfair comparison, we did not include them in the tables.

In Figures 1,2,3,4, we present qualitative results for Cityscapes → Berkeley, SYNTHIA → Berkeley, GTA5 → Berkeley and Berkeley → Cityscapes respectively. In each figure, predictions of the source-only, global alignment, and the proposed methods along with corresponding images and ground-truth segmentation maps are given. Black regions in the ground-truth maps belong to *other* class which are not evaluated at the test time. Significant differences from the baseline predictions are highlighted with white rectangular boxes.

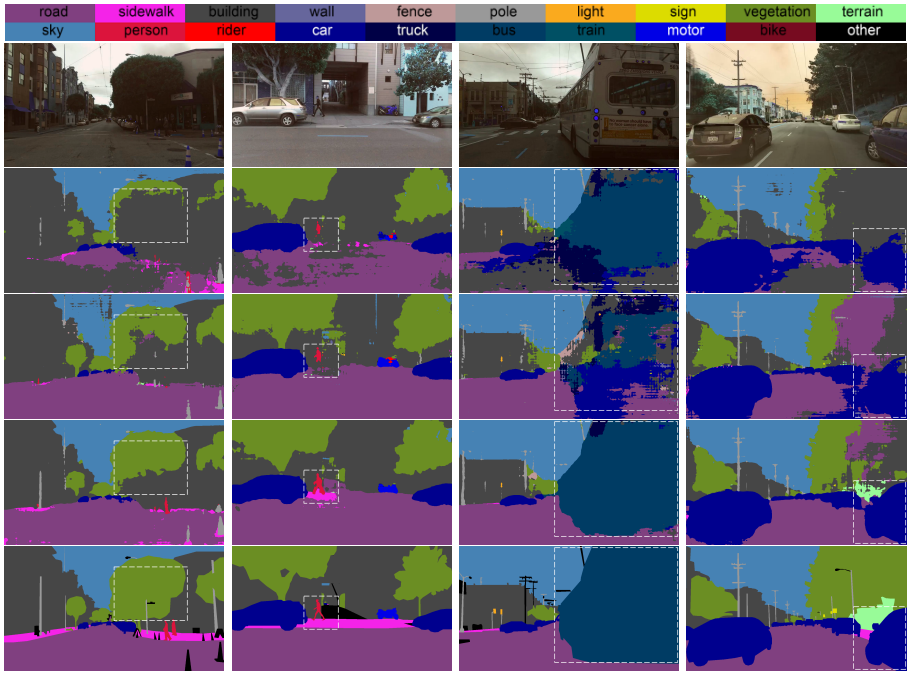

Fig. 3. GTA5 → Berkeley. From top to bottom: (1) Image, (2) source-only prediction, (3) global alignment prediction, (4) our prediction and (5) ground truth segmentation. Best viewed in color.

For Cityscapes → Berkeley, the source-only model trained on the Cityscapes over-fit the hood of the data collecting vehicle and sometimes produces erroneous segmentation for the lower part of the Berkeley images too (see Figure 1). The proposed method especially performs well on the more common classes like *car* or *building* whereas it more often fails to detect small and rare objects like *traffic signs*. These classes are challenging for compared methods too.

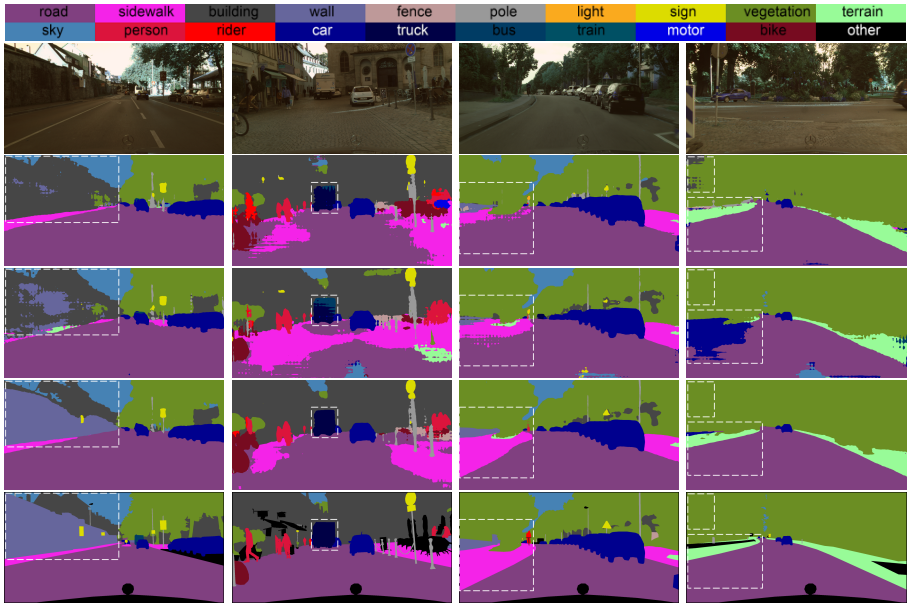

Fig. 4. Berkeley → Cityscapes. From top to bottom: (1) Image, (2) source-only prediction, (3) global alignment prediction, (4) our prediction and (5) ground truth segmentation. Best viewed in color.

As failure cases, our method performs poorly for some classes e.g. fence, train. As can be seen in Fig 5, domain shift for segmentation maps and RGB images of these classes is too large for achieving robust performance in these classes.

Fig. 5. Samples from hardest classes. Sample RGB images and binary segmentation maps for the hardest classes are given. Left two columns are for the class *fence* (hardest class for SYNTHIA → Cityscapes) and the right two columns are for *train* (hardest class for GTA5 → Cityscapes).

References

1. Yu, F., Xian, W., Chen, Y., Liu, F., Liao, M., Madhavan, V., Darrell, T.: Bdd100k: A diverse driving video database with scalable annotation tooling. arXiv preprint arXiv:1805.04687 (2018) 1
2. Hoffman, J., Wang, D., Yu, F., Darrell, T.: Fcns in the wild: Pixel-level adversarial and constraint-based adaptation. arXiv preprint arXiv:1612.02649 (2016) 2, 4
3. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. arXiv preprint arXiv:1811.12833 (2018) 3
4. Zou, Y., Yu, Z., Kumar, B., Wang, J.: Domain adaptation for semantic segmentation via class-balanced self-training. arXiv preprint arXiv:1810.07911 (2018) 4
5. Zhang, Y., Qiu, Z., Yao, T., Liu, D., Mei, T.: Fully convolutional adaptation networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. (2018) 6810–6818 4