No More Discrimination: Cross City Adaptation of Road Scene Segmenters Supplementary Material

Yi-Hsin Chen[1], Wei-Yu Chen[3,4], Yu-Ting Chen[1]*, Bo-Cheng Tsai[2]*, Yu-Chiang Frank Wang[4], Min Sun[1]

Department of {[1]Electrical Engineering,[2]Communication Engineering}, National Tsing Hua University, Taiwan
[3]Department of Electrical Engineering, National Taiwan University, Taiwan
[4]Research Center for Information Technology Innovation, Academia Sinica, Taiwan
{yhethanchen, wyharveychen, yuting2401, vigorous0503}@gmail.com
ycwang@citi.sinica.edu.tw , sunmin@ee.nthu.edu.tw

1. Visualize GA, CA and Static-Object prior

In Sec. 4.1-4.3 of the main paper, we explain how each component in our structure enhance the performance of segmentation, and also show quantitative results in experiment. Here we'll further illustrate effects of these components:

*indicates equal contribution

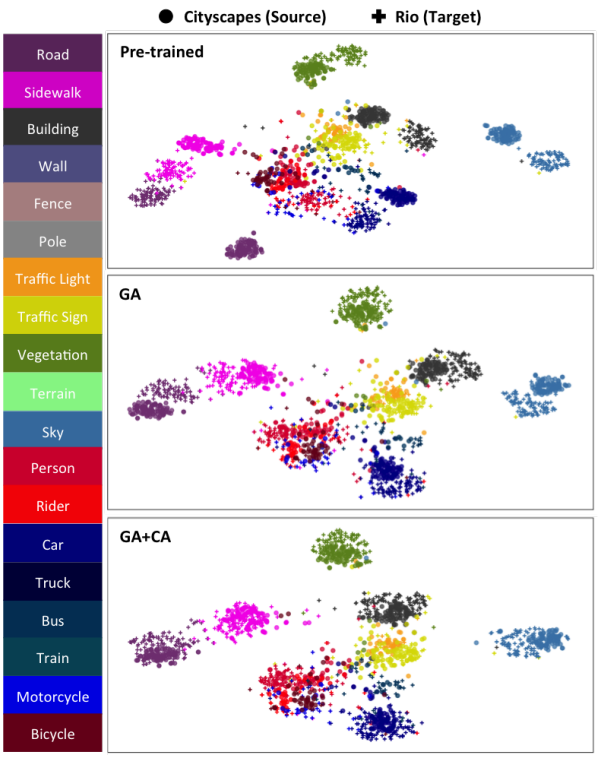

Figure 1: t-SNE visualization results. For simplicity, we only show the results of the task *Cityscapes* → *Rio*. We could clearly observe that the alignment between domains becomes better from *pre-trained* to *GA+CA*.

T-SNE Visualization To visualize the adaptation results on common feature space with t-SNE, we randomly select 100 images from each domain, and for each image we extracted its average $fc7$ feature from each class, so for both source and target we have 100 feature points from each class.

As shown in Fig. 1, with pre-trained model only, there is an obvious shift between source and target domain. After applying the global alignment (GA), the distance between clusters with same labels becomes closer, while we could still observe a gap between domains. Once we further apply the class-wise alignment (CA), the gap between domains nearly vanishes. This result again demonstrates the effectiveness of each component of our proposed method.

Harvesting Static-Object Prior In Sec. 4.3, we propose a novel pipeline to extract the static-object prior using the natural synchronization of static objects over time. For better understanding, we show some typical results of our proposed pipeline in Fig. 2. Clearly, most of the regions identified by our method truly belong to static-objects. This demonstrates the effectiveness of our method.

2. Synthetic to Real Adaptation

In Sec. 5.4 of the main paper, we have shown the quantitative results of this adaptation task in Table 3. We conclude that our method could perform well even under this challenging setting. To better support our conclusion, here we show some typical examples of this task in Fig. 3.

3. Dataset

To demonstrate the uniqueness of our dataset for road scene semantic segmenter adaptation, here we show more examples of it.

Unlabeled Image Pairs There are more examples collected at different cities with diverse appearances in Fig. 4. Valu-

Figure 2: Typical results of our static-object prior pipeline. The first row is the original unlabeled image pair of same place across time. The second row is the result of dense matching, noted by points of same color. The third row is the result of superpixel segmentation marked by different colors. Combining the results from the above two rows, we could extract static-object prior of this image pair, as shown by the red regions in the last row.

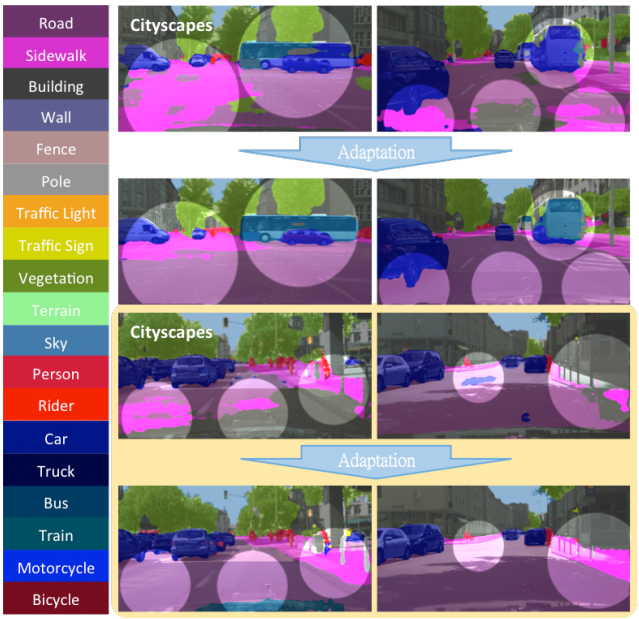

Figure 3: Examples of STNTHIA to Cityscapes adaptation task. The first/third row and second/fourth row show the results before and after adaptation, respectively. We highlight the improved regions for better visualization.

able temporal information which facilitates *unsupervised* adaptation is contained in these image pairs.

Labeled Image We also show more annotated images in Fig. 5 to demonstrate the label-quality of our dataset.

4. Label Distributions Across Cities

In Sec. 5.3 of the main paper, we conduct an experiment for cross-city adaptation. Here we show the label statistics of each city in percentage of image pixel number, which is calculated using the annotations in the testing set. As shown in Table 1, the label distributions of each city are very different. Thus, the adaptation task which we address is not trivial.

City	Frankfurt	Rome	Tokyo	Rio	Taipei
Road	36.7%	10.6%	15.2%	9.1%	11.0%
SW	6.1%	1.9%	3.2%	4.6%	1.7%
BLDG	25.8%	43.1%	37.1%	45.6%	53.0%
TL	0.2%	0.0%	0.1%	0.0%	0.1%
TS	0.6%	0.2%	0.2%	0.1%	0.3%
VEG	17.0%	21.2%	22.6%	20.4%	13.0%
Sky	4.0%	10.2%	10.6%	10.4%	10.5%
Person	1.3%	0.4%	0.6%	1.2%	0.0%
Rider	0.1%	0.1%	0.2%	0.1%	0.4%
Car	7.0%	11.6%	8.3%	6.7%	6.4%
Bus	0.8%	0.3%	1.5%	1.4%	1.2%
Motor.	0.1%	0.4%	0.1%	0.4%	1.8%
Bicycle	0.4%	0.0%	0.4%	0.1%	0.1%

Table 1: Label distribution of each city, in percentage of pixel number. SW, BLDG, TL, TS, VEG, Motor stand for Sidewalk, Building, Traffic Light, Traffic Sign, Vegetation, and Motorbike, respectively.

Figure 4: Examples of the unlabeled image pairs of different cities in our dataset. In each row, we show two image pairs at different locations in one city.

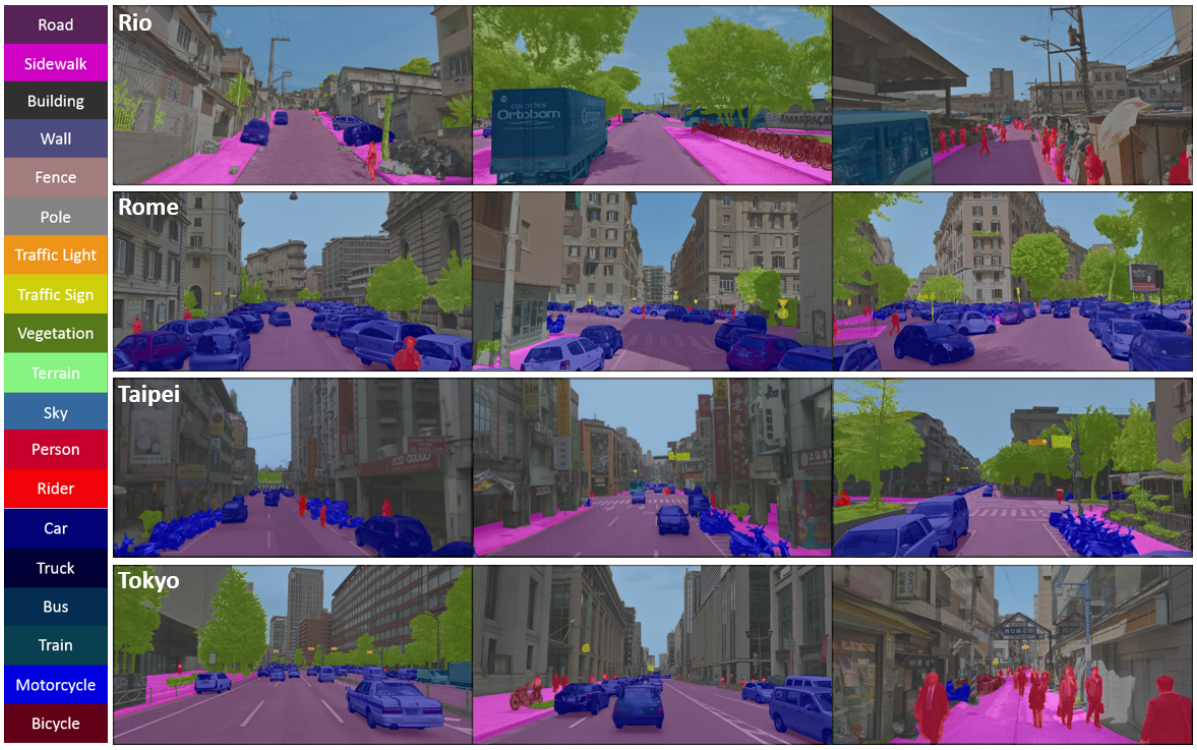

Figure 5: Examples of the labeled images of different cities in our dataset. Each image is annotated in good quality.