Domain Adaptation for Structured Output via Discriminative Patch Representations

Yi-Hsuan Tsai[1] Kihyuk Sohn[2*] Samuel Schulter[1] Manmohan Chandraker[1,3]
[1]NEC Laboratories America [2]Google Cloud [3]University of California, San Diego

1. Training Details

To train the model in an end-to-end manner, we randomly sample one image from each of the source and target domain (i.e., batch size as 1) in a training iteration. Then we follow the optimization strategy as described in Section 3.3 of the main paper. Table 1 shows the image and patch sizes during training and testing. Note that, the aspect ratio of the image is always maintained (i.e., no cropping) and then the image is down-sampled to the size as in the table.

Table 1. Image and patch sizes for training and testing.

Dataset	Cityscapes	GTA5	SYNTHIA	Oxford RobotCar
Patch size for training	32×64	36×64	38×64	-
Image size for training	512×1024	720×1280	760×1280	960×1280
Image size for testing	512×1024	-	-	960×1280

2. Relation to Entropy Minimization

Entropy minimization [1] can be used as a loss in our model to push the target feature representation F_t to one of the source clusters. To add this regularization, we replace the adversarial loss on the patch level with an entropy loss as in [3], where the entropy loss $\mathcal{L}_{en} = \sum_{u,v} \sum_k H(\sigma(F_t/\tau))^{(u,v,k)}$, H is the information entropy function, σ is the softmax function, and τ is the temperature of the softmax. The model with adding this entropy regularization achieves the IoU as 41.9%, that is lower than the proposed patch-level adversarial alignment as 43.2%. The reason is that, different from the entropy minimization approach that does not use the source distribution as the guidance, our model learns discriminative representations for the target patches by pushing them closer to the source distribution in the clustered space guided by the annotated labels.

3. More Ablation Study on Clustered Space

To validate the effectiveness of the $\mathbf{H}$ module, we conduct an experiment on GTA5-to-Cityscapes that directly computes category histograms from the segmentation output and then perform alignment. This implementation achieves an IoU 0.7% lower than our method as 41.3%. A possible reason is that we use the $\mathbf{H}$ module that involves learnable parameters to estimate K-means memberships, whereas directly computing category histograms would solely rely on updating the segmentation network $\mathbf{G}$, which causes more training difficulty.

4. More Details on Pseudo Label Re-training

We use the official implementation of [2] provided by the authors. In this case, we consider our target samples as unlabeled data used in [2] under the semi-supervised setting. The same discriminator in the output space and the loss function are then adopted as in [2].

*The work was done at NEC Laboratories America.

5. Result of Adapting Cityscapes to Oxford RobotCar

In Table 2, we present the complete result for adapting Cityscapes (sunny condition) to Oxford RobotCar (rainy scene). We compare the proposed method with the model without adaptation and the output space adaptation approach [4]. More qualitative results are provided in the supplementary video.

Table 2. Results of adapting Cityscapes to Oxford RobotCar.

	Cityscapes → Oxford RobotCar									
Method	road	sidewalk	building	light	sign	sky	person	automobile	two-wheel	mIoU
Without Adaptation	79.2	49.3	73.1	55.6	37.3	36.1	54.0	81.3	49.7	61.9
Output Space [4]	95.1	64.0	75.7	61.3	35.5	63.9	58.1	84.6	57.0	69.5
Ours	94.4	63.5	82.0	61.3	36.0	76.4	61.0	86.5	58.6	72.0

6. Qualitative Comparisons

We provide more visual comparisons for GTA5-to-Cityscapes and SYNTHIA-to-Cityscapes scenarios from Figure 1 to Figure 3. In each row, we present the results of the model without adaptation, output space adaptation [4], and the proposed method. We show that our approach often yields better segmentation outputs with more details and produces less noisy regions.

References

[1] Y. Grandvalet and Y. Bengio. Semi-supervised learning by entropy minimization. In *NIPS*, 2004. 1

[2] Wei-Chih Hung, Yi-Hsuan Tsai, Yan-Ting Liou, Yen-Yu Lin, and Ming-Hsuan Yang. Adversarial learning for semi-supervised semantic segmentation. In *BMVC*, 2018. 2

[3] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Unsupervised domain adaptation with residual transfer networks. In *NIPS*, 2016. 1

[4] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *CVPR*, 2018. 2, 3, 4, 5

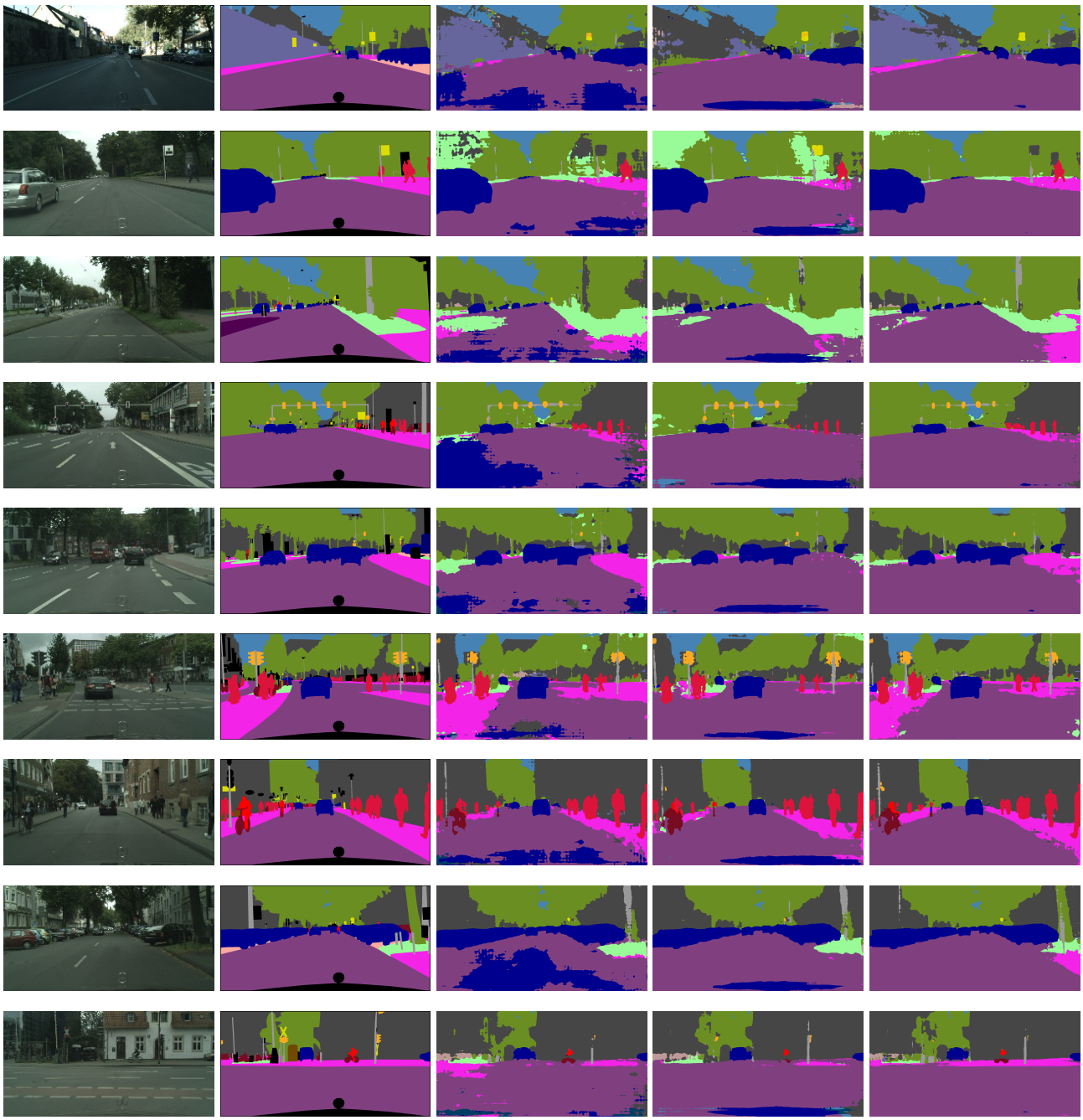

Figure 1. Example results of adapted segmentation for the GTA5-to-Cityscapes setting. For each target image, we show results before adaptation, output space adaptation [4], and the proposed method.

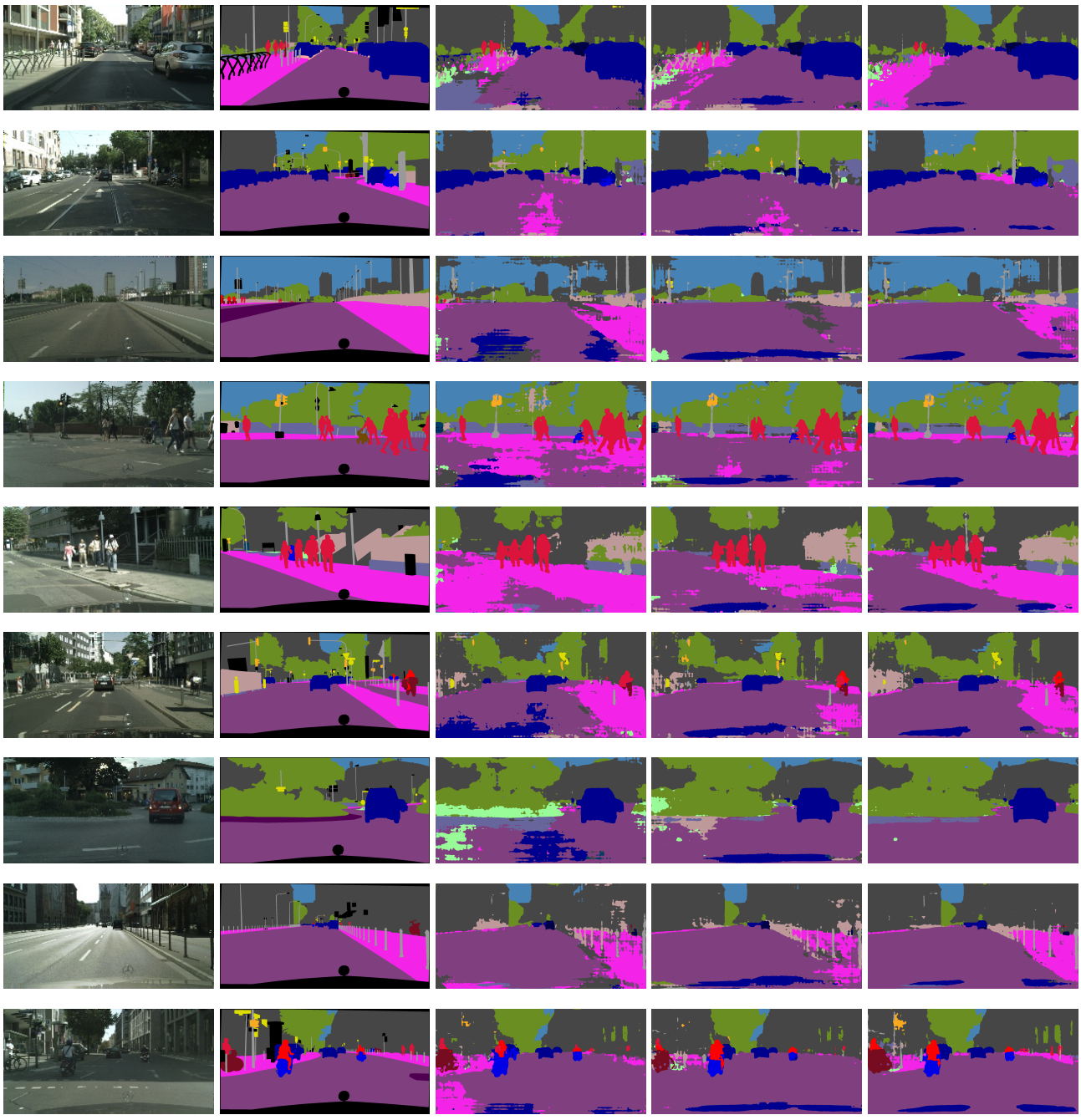

Figure 2. Example results of adapted segmentation for the GTA5-to-Cityscapes setting. For each target image, we show results before adaptation, output space adaptation [4], and the proposed method.

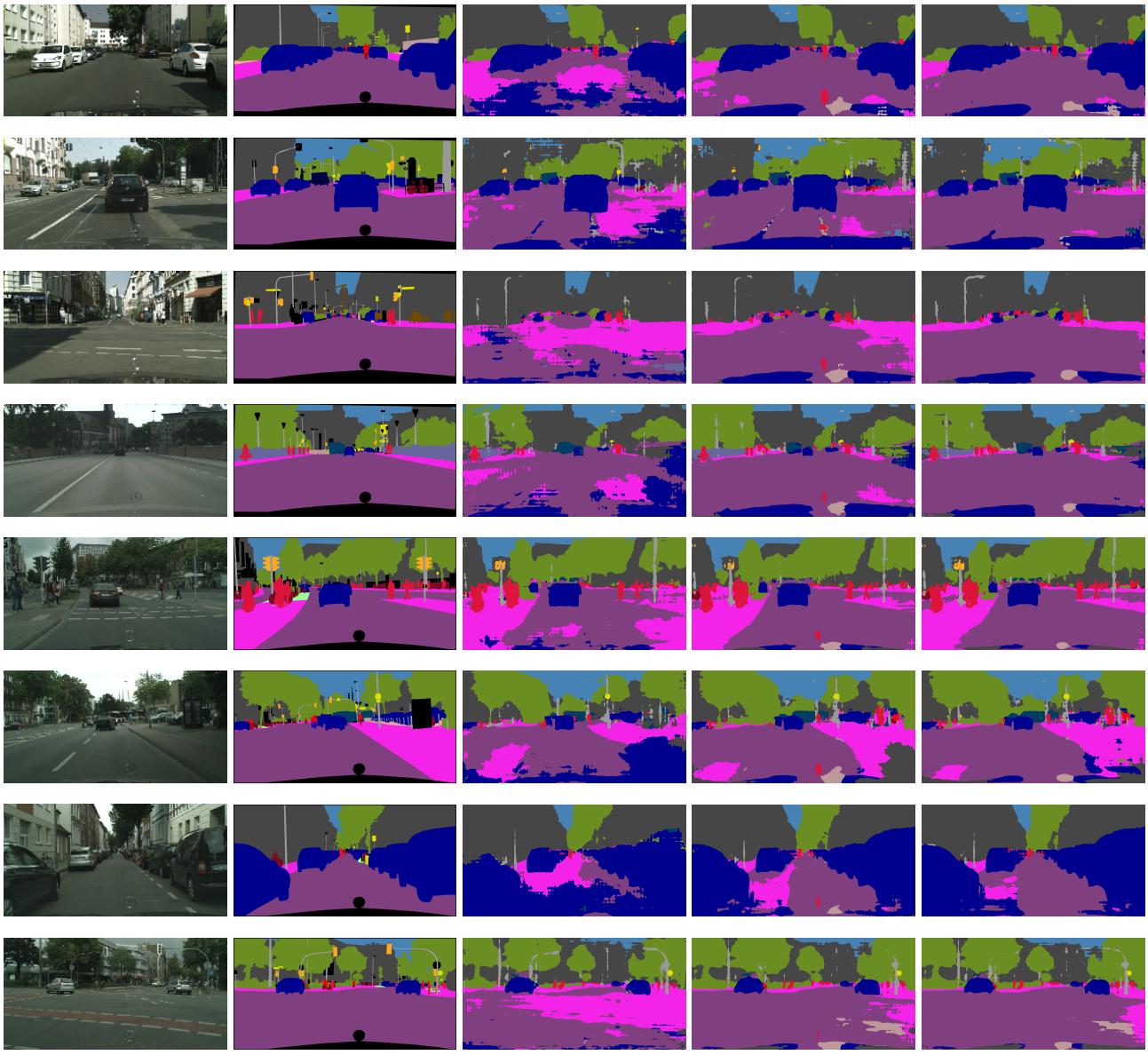

Figure 3. Example results of adapted segmentation for the SYNTHIA-to-Cityscapes setting. For each target image, we show results before adaptation, output space adaptation [4], and the proposed method.