Domain Adaptive Semantic Segmentation Using Weak Labels

Sujoy Paul[1], Yi-Hsuan Tsai[2], Samuel Schulter[2],
Amit K. Roy-Chowdhury[1], Manmohan Chandraker[2,3]

[1]UC Riverside [2]NEC Labs America [3]UC San Diego

1 Overview

In this supplementary material, we provide more results of ablation study and using oracle-weak labels, including image-level and point supervisions. In addition, we present parameter analysis on GTA5 → Cityscapes, when using 1) pseudo-weak labels, and 2) image-level oracle-weak labels. Moreover, we show more visual results of 1) category-wise probability, and 2) final semantic segmentation comparisons. Finally, we also empirically analyze the performance of our model on a different architecture and a different dataset.

2 Ablation Study

We provide an ablation study extended from Table 1 and 2 of the main paper. In Table 1 and 2, we add the factor with or without using the pixel-level adaptation (denoted as PA). From the results, we show that our proposed weak label loss ($\mathcal{L}_c$) and category-wise alignment loss ($\mathcal{L}_{adv}^C$) are both complementary to the PA module, under both the case of UDA and WDA settings.

Table 1: Ablation of the proposed loss functions for GTA5→Cityscapes.

	Supervision	$\mathcal{L}_c$	$\mathcal{L}_{adv}^C$	PA	mIoU
UDA	No Adapt.				36.6
UDA	Baseline [4]				41.4
UDA	Pseudo-Weak	✓			44.2
UDA	Pseudo-Weak	✓	✓		45.6
UDA	Pseudo-Weak	✓		✓	46.7
UDA	Pseudo-Weak	✓	✓	✓	**48.2**
WDA	Oracle-Weak	✓			50.8
WDA	Oracle-Weak	✓	✓		52.1
WDA	Oracle-Weak	✓		✓	52.0
WDA	Oracle-Weak	✓	✓	✓	**53.0**

Table 2: Ablation of the proposed loss functions for SYNTHIA→Cityscapes.

	Supervision	$\mathcal{L}_c$	$\mathcal{L}_{adv}^C$	PA	mIoU	mIoU*
UDA	No Adapt.				33.5	38.6
UDA	Baseline [4]				39.5	45.9
UDA	Pseudo-Weak	✓			41.7	49.0
UDA	Pseudo-Weak	✓	✓		42.7	49.9
UDA	Pseudo-Weak	✓		✓	43.0	50.6
UDA	Pseudo-Weak	✓	✓	✓	**44.3**	**51.9**
WDA	Oracle-Weak	✓			47.8	56.0
WDA	Oracle-Weak	✓	✓		49.2	57.2
WDA	Oracle-Weak	✓		✓	49.8	57.8
WDA	Oracle-Weak	✓	✓	✓	**50.6**	**58.5**

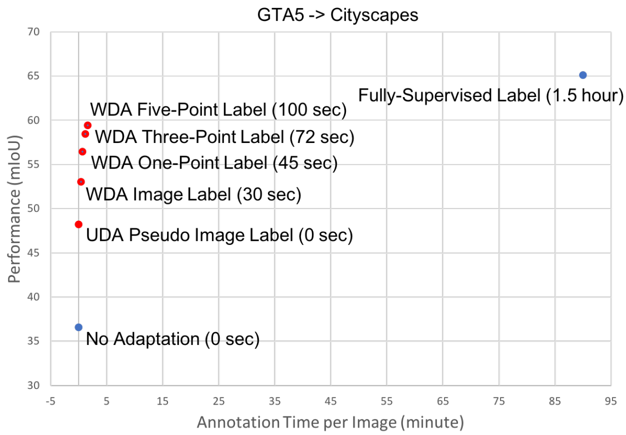


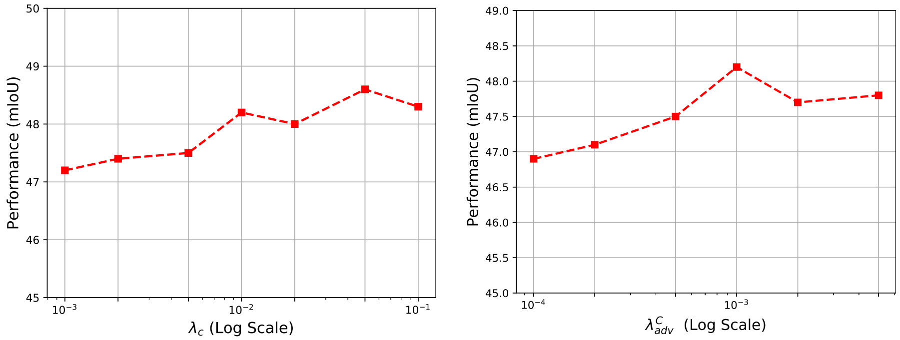
Fig. 1: Performance comparison on GTA5 → Cityscapes with different levels of supervision on the target domain.

3 Performance for Oracle-weak Labels

To further investigate the effectiveness of applying our framework for using oracle-weak labels, we extend the setting of point supervision from annotating 1 point to more points. In Fig. 1, we compare the performance v.s. annotation cost on GTA5 → Cityscapes, and show that with a small amount of increase in annotation time, the performance can be improved using more points. For instance, when using 5-point supervision (100 seconds per image), the mIoU reaches 59.4%, which is close to the fully-supervised setting with 65.1% mIoU but requiring significantly longer annotation process (1.5 hours per image). This also demonstrates the usefulness of the proposed novel WDA setting and our framework that can take different types of oracle-weak labels to improve the performance. Note that in experiments, weak labels are extracted from ground truths provided in the dataset. We estimate the annotation time by averaging the time required for a human annotator to label a portion of the dataset.

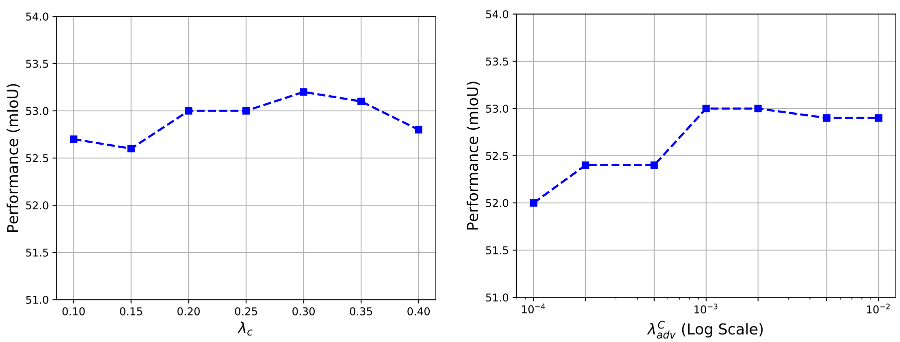
4 Parameter Analysis for Pseudo-Weak Labels

Fig. 2 presents two plots for the parameter analysis when using pseudo-weak labels. In Fig. 2(a), we fix $\lambda^{C}_{adv} = 0.001$ and show that our model achieves the mIoU larger than 47.5% under a range of $\lambda_c = [0.005, 0.1]$. When fixing $\lambda_c = 0.01$, Fig. 2(b) shows that the model performs well under a range of $\lambda^{C}_{adv} = [0.0005, 0.005]$. However, when we increase λ^{C}_{adv} to be larger than 0.01, the adversarial training process may become unstable and decreases the performance to 46.2%. In addition, decreasing λ^{C}_{adv} would give less focus on alignment and gradually degrades the performance, which shows the importance of our alignment process.



(a) Weight on Classification Loss (b) Weight on Category-wise Alignment Loss

Fig. 2: Plots presenting the hyper-parameter analysis of the parameters λ_c on the classification loss using pseudo-weak labels and λ^C_{adv} on the category-wise alignment loss.



(a) Weight on Classification Loss (b) Weight on Category-wise Alignment Loss

Fig. 3: Plots presenting the hyper-parameter analysis of the parameters λ_c on the classification loss using oracle-weak labels and λ^C_{adv} on the category-wise alignment loss.

5 Parameter Analysis for Oracle-Weak Labels

Fig. 3 presents two plots for the parameter analysis when using image-level oracle-weak labels. Since such weak labels are accurate, Fig. 3(a) (fixing $\lambda_{adv} = 0.001$) shows that the performance is quite stable when changing λ_c. Moreover, in Fig. 3(b) fixing $\lambda_c = 0.2$, the performance starts to drop when decreasing λ_{adv} as the alignment process becomes weaker.

6 Category-wise Visualization

Fig. 5 presents some example visualizations showing the category-wise spatial probabilities before using any weak labels for adaptation, after using pseudo-weak labels and finally after using oracle-weak labels for adaptation. For example, before using weak labels (e.g., *Sidewalk* and *Fence*), some regions are highlighted incorrectly, and those regions are eliminated after using pseudo-weak labels (UDA) and oracle-weak labels (WDA). In addition, even with a small portion being highlighted before using weak labels (e.g., *Sign*, *Rider*, *Bus*), the probability maps become more prominent after using weak labels. Moreover, we present two failure cases of the pseudo-weak labels in Fig. 4. In the first row, category *Fence* occurs, but is not detected in the pseudo-weak label. In the second row, category *Wall* does not appear, but is detected in the pseudo-weak label.

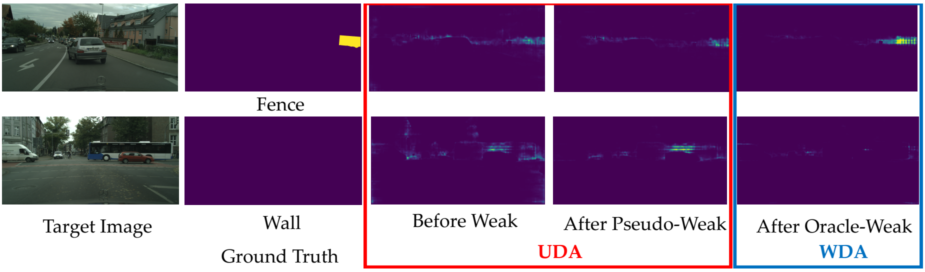
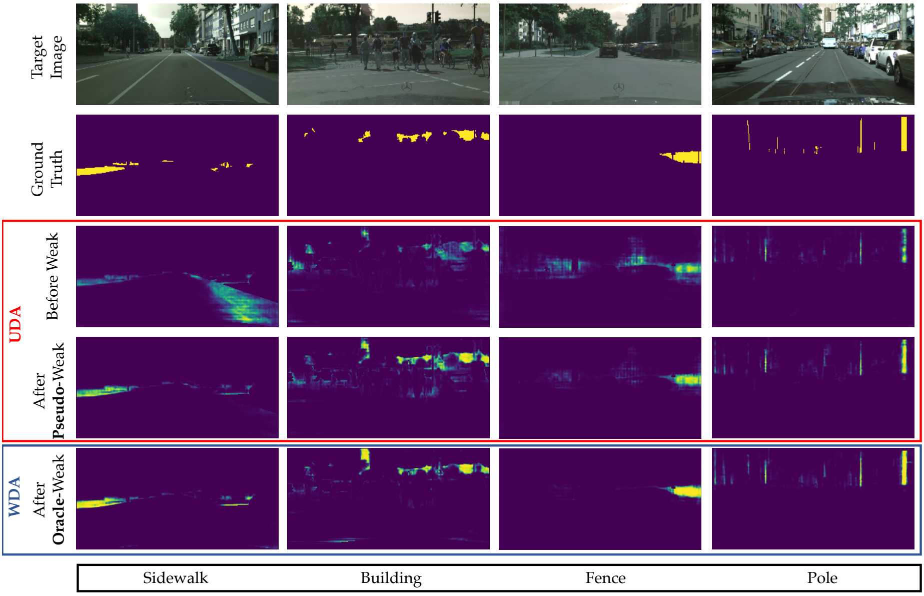


Fig. 4: Visualizations for two types of failure cases.



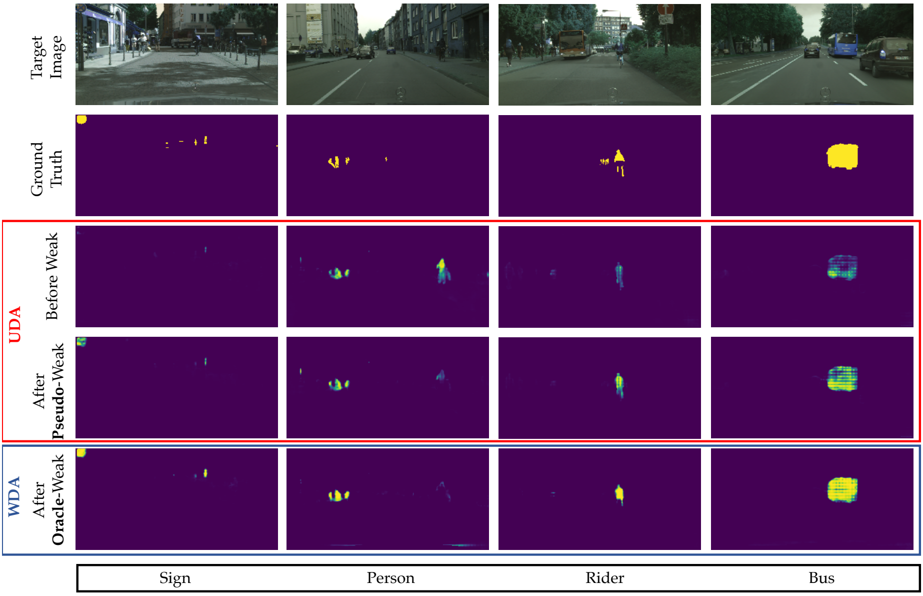


Fig. 5: Visualizations of category-wise segmentation prediction probability before and after using the pseudo-weak labels on GTA5 → Cityscapes. Before adaptation, the network only highlights the areas partially with low probability, while using the pseudo-weak labels helps the adapted model obtain much better segments, and is closer to the model using oracle-weak labels.

7 Semantic Segmentation Visualization

Fig. 6 presents the semantic segmentation results before and after using weak labels for adaptation. The UDA method without using any weak labels produces more erroneous results in some portions and may miss some of the categories within a small area, such as sign, pole, etc. However, using the pseudo-weak labels enhances the segmentation and helps our model better identify the categories which originally have a lower confidence. Moreover, using oracle-weak labels is able to further improve the segmentation performance.

8 More Results on Architecture and Dataset

Table 3 presents segmentation performance using the VGG16 architecture with GTA5 as source and Cityscapes as target. Our method performs better than other UDA methods. We also present results for the WDA case with oracle-weak labels, i.e., image or point labels, which produces higher performance than the UDA methods.

Moreover, we test our method with GTA5 as source and Foggy Cityscapes [3] as target. There is a parameter to choose the level of fog in the images, and we set that to 0.02 in our experiments. The results are presented in Table 4. We can observe consistent improvements as in other datasets.

Table 3: Results of adapting GTA5 to Cityscapes with VGG16. The top group is for Unsupervised Domain Adaptation (UDA), while the bottom group presents our method's performance using the oracle-weak labels for Weakly-supervised Domain Adaptation (WDA) that use either image-level or point supervision.

	GTA5 → Cityscapes																			
Method	road	sidewalk	building	wall	fence	pole	light	sign	veg	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mIoU
AdaptOutput [4]	87.3	29.8	78.6	21.1	18.2	22.5	21.5	11.0	79.7	29.6	71.3	46.8	6.5	80.1	23.0	26.9	0.0	10.6	0.3	35.0
AdvEnt [6]	86.9	28.7	78.7	28.5	**25.2**	17.1	20.3	10.9	80.0	26.4	70.2	47.1	8.4	81.5	26.0	17.2	**18.9**	11.7	1.6	36.1
CLAN [2]	88.0	30.6	79.2	23.4	20.5	**26.1**	23.0	**14.8**	81.6	**34.5**	72.0	45.8	7.9	80.5	**26.6**	**29.9**	0.0	10.7	0.0	36.6
SSF-DAN [1]	**88.7**	32.1	**79.5**	29.9	22.0	23.8	21.7	10.7	80.8	29.8	**72.5**	49.5	16.1	**82.1**	23.2	18.1	3.5	**24.4**	8.1	37.7
AdaptPatch [5]	87.3	**35.7**	**79.5**	**32.0**	14.5	21.5	24.8	13.7	80.4	32.0	70.5	50.5	16.9	81.0	20.8	28.1	4.1	15.5	4.1	37.5
Ours (UDA)	87.1	**35.7**	78.6	24.9	22.7	21.8	**26.5**	11.7	**82.1**	32.1	70.4	**50.6**	**18.3**	77.4	21.7	24.6	7.6	16.3	**19.3**	**38.4**
Ours (Image)	88.0	46.8	81.6	22.3	35.2	27.4	29.2	27.0	82.4	35.4	80.7	57.1	29.0	83.2	38.0	56.4	23.3	29.8	5.5	46.2
Ours (Point)	93.6	62.7	81.4	29.6	33.7	30.7	29.7	38.2	81.5	43.0	81.7	54.3	28.8	83.8	42.9	52.5	38.4	27.1	49.8	51.8

Table 4: Results of adapting GTA5 to Foggy Cityscapes with ResNet101. The top group is for Unsupervised Domain Adaptation (UDA), while the bottom group presents our method's performance using the oracle-weak labels for Weakly-supervised Domain Adaptation (WDA) that use either image-level or point supervision.

	GTA5 → Cityscapes																			
Method	road	sidewalk	building	wall	fence	pole	light	sign	veg	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	mIoU
No Adapt.	78.8	11.8	67.8	15.1	15.6	19.5	20.6	12.1	63.6	19.3	60.3	49.3	22.6	55.6	17.2	14.9	0.0	19.2	27.0	31.0
AdaptOutput [4]	87.3	24.9	70.2	15.4	18.7	19.6	24.9	18.6	**69.3**	28.2	64.4	49.5	24.1	74.0	**17.6**	21.2	2.1	**27.5**	35.9	36.5
Ours (UDA)	**88.8**	**27.8**	**71.0**	**21.7**	**21.8**	**26.4**	**33.1**	**26.2**	68.7	**29.4**	**66.3**	**55.4**	**27.2**	**77.1**	11.8	**24.0**	**5.7**	14.7	**39.3**	**38.8**
Ours (Image)	89.0	32.8	76.5	22.0	26.5	29.8	35.3	34.8	77.4	32.8	71.7	60.1	35.0	84.7	33.6	42.0	19.0	30.8	44.1	46.2
Ours (Point)	92.7	55.0	80.0	28.3	29.3	34.2	37.4	45.8	79.9	32.8	73.4	62.4	34.0	85.8	37.2	50.6	19.3	28.1	53.7	50.5

References

1. Du, L., Tan, J., Yang, H., Feng, J., Xue, X., Zheng, Q., Ye, X., Zhang, X.: Ssf-dan: Separated semantic feature based domain adaptation network for semantic segmentation. In: ICCV (2019) 6
2. Luo, Y., Zheng, L., Guan, T., Yu, J., Yang, Y.: Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In: CVPR (2019) 6
3. Sakaridis, C., Dai, D., Van Gool, L.: Semantic foggy scene understanding with synthetic data. IJCV **126**(9), 973–992 (2018) 6
4. Tsai, Y.H., Hung, W.C., Schulter, S., Sohn, K., Yang, M.H., Chandraker, M.: Learning to adapt structured output space for semantic segmentation. In: CVPR (2018) 1, 6, 7
5. Tsai, Y.H., Sohn, K., Schulter, S., Chandraker, M.: Domain adaptation for structured output via discriminative patch representations. In: ICCV (2019) 6
6. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: CVPR (2019) 6

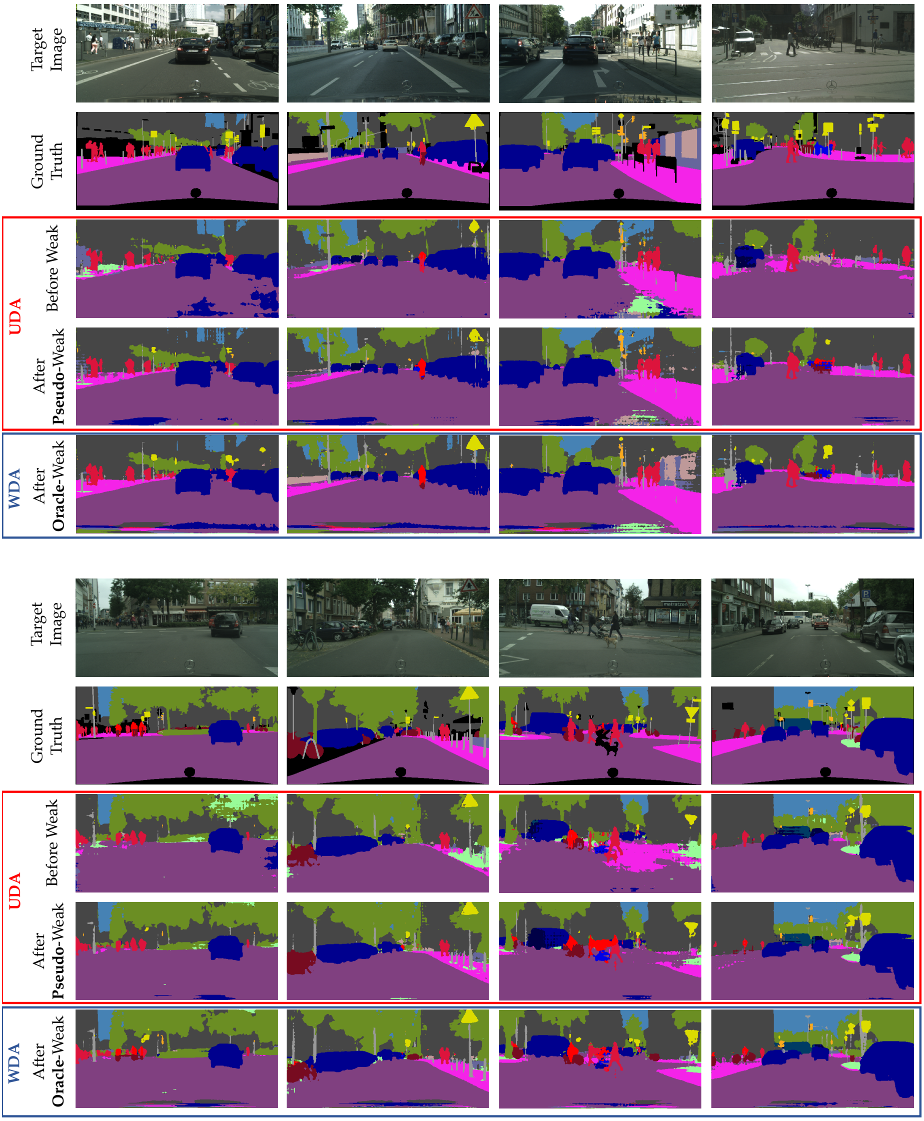


Fig. 6: Example results of adapted segmentation for GTA5 → Cityscapes with and without using weak labels for adaptation. The visualizations show that using pseudo-weak labels, the segmentation become more structured and some of the categories are better segmented. Using oracle-weak labels further improves the segmentation quality.