TACS: Taxonomy Adaptive Cross-Domain Semantic Segmentation — Supplementary Material —

Rui Gong[1], Martin Danelljan[1], Dengxin Dai[3], Danda Pani Paudel[1], Ajad Chhatkuli[1], Fisher Yu[1], and Luc Van Gool[1,2]

[1] Computer Vision Lab, ETH Zurich, Switzerland
{gongr,martin.danelljan,paudel,ajad.chhatkuli,vangool}@vision.ee.ethz.ch
i@yf.io
[2] VISICS, ESAT/PSI, KU Leuven, Belgium
[3] VAS, MPI for Informatics, Germany ddai@mpi-inf.mpg.de

In this supplementary material, we provide the additional information for,

S1 clarification on the code implementation, and discussion on the future work, societal impact and limitations,
S2 comparison between our proposed TACS and other domain adaptation problems,
S3 detailed implementation of our proposed framework,
S4 detailed information of involved datasets in our experiments,
S5 additional quantitative and qualitative experimental results.

S1 Code, Future Work, Societal Impact and Limitations

Code Implementation. Our implementation is publicly available at https://github.com/ETHRuiGong/TADA.
Future Work. Inspired by [11, 14, 20], the integration of our method and the active learning based methods [11, 14, 20], identifying the informative instances to label under TACS setting, serves as a promising future work direction.
Potential Negative Societal Impact and Limitations. *Societal Impact:* Our proposed approach provides the potential to adapt the semantic segmentation model even under the inconsistent taxonomy, saving much cost and effort for labeling when new data and new requirements come. Thus, there is also a risk for reduced need of data labelling, leading fewer jobs in this domain and potential unemployment. *Limitations:* The main limitation is that the domain adaptation approach is yet to achieve the performance of fully supervised training.

S2 Comparison with Other Domain Adaptation Problems

In Sec. 2 of the main paper, we compare different domain adaptation problems with our newly proposed taxonomy adaptive cross-domain semantic segmentation (TACS) problem. Here we provide more clarification. We firstly clarify the difference overview between TACS and other domain adaptation problems, and then explain the detailed difference between TACS and different domain adaptation problems, respectively.

Overview. As discussed in the abstract and Sec. 1 of the main paper, our TACS tackles the inconsistent taxonomy between the source and target domain, motivated by the fact that the target domain task requires a different taxonomy than the one imposed by the source domain in many real world settings, due to different scenarios, different granularity levels classes, inconsistent annotation practices, or the strive towards an increasingly fine-grained taxonomy [7]. Our TACS includes three typical inconsistent taxonomy types, *i.e.,* i) open taxonomy, ii) coarse-to-fine taxonomy and iii) implicitly-overlapping taxonomy. We then compare our TACS with other domain adaptation problems. The traditional unsupervised domain adaptation (UDA) [23,24], partial domain adaptation (PDA) [2] and few-shot/semi-supervised domain adaptation (FS/SS DA) [10,21,30] all assume the consistent taxonomy (see Fig. 1 of the main paper) between the source domain and the target domain, while ignoring the inconsistent taxonomy between the source domain and the target domain tackled by our TACS. The open-set (OS) [12,19]/ universal (US) [29] / zero-shot (ZS) [1]/ class-incremental (CI) [6] domain adaptation problems touch upon the specific or special case of the open taxonomy setting in our TACS. However, the open taxonomy setting in our TACS provides a more flexible and practical setting (see next analysis in TACS *vs.* OS/US DA and TACS *vs.* ZS/CI DA for more details), compared with OS/US/ZS/CI DA. Besides, the coarse-to-fine taxonomy and implicitly-overlapping taxonomy settings in our TACS are not touched upon by above other domain adaptation problems. Thus, our TACS provides a more general, flexible and practical setting, allowing for different types of inconsistent taxonomies, *e.g.,* different granularity levels classes, between the source and the target domain. Next, we compare our TACS with different domain adaptation problems in more detail, respectively.

TACS *vs.* **Unsupervised Domain Adaptation (UDA).** The traditional UDA [23,24] only focuses on the image-level domain gap, but ignores the label-level domain gap (cf. Fig. 1 of the main paper), *i.e.,* assuming the consistent taxonomy between the source domain and the target domain.

TACS *vs.* **Partial Domain Adaptation (PDA).** The implicitly-overlapping taxonomy in our TACS is totally different from PDA [2]. PDA only assumes the reduced label space from the source domain to the target domain, *e.g.,* {"vehicle", "bicycle"} → {"bicycle"}, which actually still assumes consistent taxonomy between the source domain and the target domain (cf. (c) in Fig. 1 of the main paper). However, the implicitly-overlapping taxonomy setting in our TACS touches the problem that, for a certain class in the source domain, one or more of its sub-classes are merged into other classes in the target domain, *e.g.,* {"vehicle", "bicycle"} → {"car", "cycle"}, which tackles the inconsistent taxonomy between the source domain and the target domain (cf. (f) in Fig. 1 of the main paper).

TACS *vs.* **Few-Shot/Semi-Supervised Domain Adaptation (FS/SS DA).** FS/SS DA [10,21,30] aims at improving the domain adaptation performance by introducing the few-shot fully labeled target domain samples. However, FS/SS DA still assumes the consistent taxonomy between the source domain and the target domain.

TACS *vs.* Open-Set/Universal Domain Adaptation (OS/US DA). OS/US DA [12, 19, 29] aims at recognizing the new unseen classes in the target domain together as an "unknown" class, which can be seen as a special case of our open taxonomy setting in our TACS. Differently, the open taxonomy setting in our TACS aims at recognizing different new unseen classes explicitly and separately. For example, assuming {"terrain", "train"} are the new unseen classes in the target domain, OS/US DA just aims at recognizing the pixels of {"terrain", "train"} classes as the "unknown" class pixel together. However, the open taxonomy setting in our TACS aims at recognizing the pixels of {"terrain", "train"} classes as the "terrain" and "train" classes explicitly and seperately, as the recognition of the seen class. Besides, OS/US DA does not consider the coarse-to-fine taxonomy and implicitly-overlapping taxonomy setting in our TACS.

TACS *vs.* Zero-Shot/Class-Incremental Domain Adaptation (ZS/CI DA). Similar to the open taxonomy setting of our TACS, ZS/CI DA [1,6] aims at recognizing the new unseen classes in the target domain explicitly and separately, which can be seen as a specific case of the open taxonomy setting of our TACS. However, ZS/CI DA only considers the case where the unseen classes are absent in the source domain. In contrast, the open taxonomy setting in our TACS also allows for the unseen classes to exist in the source domain, where they are unlabelled. Besides, ZS/CI DA does not consider the coarse-to-fine taxonomy and implicitly-overlapping taxonomy setting in our TACS.

S3 Framework Implementation

In the main paper, we propose the new taxonomy adaptive cross-domain semantic segmentation (TACS) problem, which allows inconsistent taxonomies between the source domain and the target domain in the domain adaptation for semantic segmentation. TACS approaches the domain adaptation for semantic segmentation on both of the image level and the label level. In order to address the TACS problem, a set of pseudo-labelling techniques and the contrastive learning scheme are developed to reduce both of the label-level and image-level domain gap (cf. Sec. 3 of the main paper). Our proposed complete approach demonstrates the strong performance under different TACS settings, open taxonomy, coarse-to-fine taxonomy and implicitly-overlapping taxonomy (cf. Table 2, Table 3 and Table 4 of the main paper). Moreover, our suggested mixed-sampling and contrastive learning based scheme outperforms the state-of-the-art (SOTA) methods by a large margin, under traditional unsupervised domain adaptation (UDA) setting (cf. Table 1 of the main paper). Here we present the implementation details of our proposed framework.

Batch Size. For the open taxonomy, coarse-to-fine taxonomy and implicitly-overlapping taxonomy experiments of TACS in Sec. 4 of the main paper, in each training batch, there are 2 source domain images, 2 unlabeled target domain images and 2 few-shot labeled target domain images mixed in the bilateral mixed sampling module. For the consistent taxonomy experiments of UDA in Sec. 4 of the main paper, we strictly follow the traditional UDA setting, and the target domain is completely unlabelled. Therefore, under UDA setting, in each training batch, there are 2 source domain images and 2 unlabelled target domain images mixed in the class mixed sampling way [22].

Parameters. The source domain images are resized to 1280×720, and the target domain images are resized to 1024×512. And the random crop with size 512×512 is then adopted. We adopt the SGD optimizer to train the semantic segmentation network, whose momentum is set as 0.9 and the weight decay is set to 5×10^{-4}. The learning rate is set as 2.5×10^{-4}, with polynomial decay of power 0.9. The iteration T in Sec. 3.5 for starting training the RL module is set as 130000. The total training iteration is set as 250000.

Contrastive Learning. We adopt the 2048-dim output vector of the final layer of feature extractor, *i.e.,* , the layer before the classifier, of the Deeplab-v2 framework. The 2048-dim vector is mapped to a 256-dim vector with a projection head, composed of 1x1 Conv, Batchnorm, ReLU, 1x1 Conv layers. The 256-dim vector is then adopted as the pixel-wise feature. For each mini-batch, we use 100 anchor pixel samples per category. The 100 pixel samples of the same category are taken as positive samples, while the other pixel samples of different categories are all taken as negative samples.

Baseline Setup. In the baseline methods setup of Table 2, Table 3 and Table 4 in the main paper, we add the additional supervised loss to train the model in the supervised way, with the few-shot/partially labeled samples in the target domain. For the baseline methods which adopt the pseudo-label based training strategy, such as FDA [28], IAST [9], and DACS [22], the few-shot/partial label on the target domain samples is combined with the generated pseudo-label to attain the final pseudo-label. I.e., in the pseudo-label generation process on the few-shot/partially labeled samples, we adopt the ground-truth label for the labeled parts, while we adopt the generated pseudo-label for other unlabeled parts.

Compute Resources. The code is implemented with PyTorch [13]. Experiments are conducted on an NVIDIA GeForce RTX 2080 Ti GPU, with 11GB memory, where it takes 3 days for training the whole 250000 iterations. In the whole investigation process of our paper, the total compute used is around 390×3 GPU days.

S4 Datasets Information

As introduced in Sec. 4 of the main paper, there are 4 datasets in total involved in our experiments, including SYNTHIA [16], GTA5 [15], Synscapes [27] and Cityscapes [4]. Here we provide more information about the datasets.

SYNTHIA. SYNTHIA is a syntheic image dataset, consisting of photo-realistic images rendered from a virtual city. We adopt SYNTHIA-RAND-CITYSCAPES subset, including 9400 densely labeled synthetic images. SYNTHIA is licensed under a CC BY-NC-SA 3.0 license.

GTA5. GTA5 is a synthetic image dataset, containing 24966 urban scene images. The images in GTA5 dataset are rendered from game engine, and densely labeled with pixel-level semantic annotation. The scene of GTA5 dataset is based on the city of Los Angeles. We were unable to find the license for the GTA5 dataset. But the code for extracting the GTA5 dataset image from the game engine is released under the MIT license.
Synscapes. Synscapes is a photo-realistic synthetic dataset, created with physically based rendering techniques. Synscapes is built for street scene parsing, composed of 25000 densely pixel-level annotated images. Synscapes customizes the license, *i.e.,* Synscapes grants a non-exclusive, non-transferable, non-sublicensable, worldwide license to use the dataset for non-commercial purposes.
Cityscapes. Cityscapes is a real street scene image dataset, collected from different European cities. We adopt the training set of Cityscapes during the training stage, covering 2975 images. And we use the validation set of Cityscapes, including 500 images, to evaluate the performance of the semantic segmentation model. Cityscapes customizes the license, *i.e.,* Cityscapes is made freely available to academic and non-academic entities for non-commercial purposes such as academic research, teaching, scientific publications, or personal experimentation.
Whether the datasets cover personally identifiable information or offensive content? The SYNTHIA, GTA5 and Synscapes are all synthetic image datasets, and are rendered from the virtual city or game engine. The personally identifiable information or offensive content is not found in them. Cityscapes is a real street scene image dataset, but Cityscapes is for non-commercial use only. Even though Cityscapes covers the "person" class as one of the semantic annotation classes, the personally identificable information or offensive content is also not found in Cityscapes. Besides, Citysacpes creators state that, if any people find themselves or their personal belongings in the data, they will immediately remove the respective images from their servers after receiving the contact from the people.

S5 Additional Experimental Results

In Sec. 4 of the main paper, we report the experimental results under the traditional UDA setting and different TACS settings, *i.e.,* open taxonomy, coarse-to-fine taxonomy and implicitly-overlapping taxonomy. Here we provide additional quantitative and qualitative experimental results to further prove the effectiveness of our proposed approach.

S5.1 TACS: Coarse-to-Fine Taxonomy involving More Classes

In order to prove the effectiveness of our proposed approach when dealing with the inconsistent taxonomy involving more classes, we provide the experimental results under the coarse-to-fine taxonomy setting, with more fine-grained classes in the target domain.

Table S1: Coarse-to-Fine Taxonomy: GTA5→Cityscapes. The “moving object” class in the GTA5 dataset is fine-grained into 8 classes in the Cityscapes dataset. The gray columns are the 8 fine-grained classes in the Cityscapes and corresponding mean IoU of these classes.

Method	Road	SW	Build	Wall	Fence	Pole	TL	TS	Veg	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	MC	Bike	mIoU	mIoU
Source	71.59	20.93	67.54	10.00	15.49	24.15	29.90	19.46	79.83	19.10	74.07	34.95	10.53	67.43	9.98	17.72	7.86	4.75	25.14	22.30	32.13
IAST [9]	81.87	35.74	79.58	37.35	25.77	32.26	**45.14**	39.14	85.34	34.09	85.14	57.58	27.32	81.64	28.01	45.54	26.03	21.58	44.28	41.50	48.08
Ours	**95.35**	**68.30**	**86.75**	**41.39**	**38.95**	**36.62**	43.96	**49.49**	**87.64**	**45.90**	**87.43**	**63.96**	**28.31**	**88.41**	**45.41**	**59.17**	**57.34**	**37.02**	**57.13**	**54.59**	**58.87**

Table S2: Consistent Taxonomy: GTA5→Cityscapes. The mIoU is over 19 classes. In the UDA setting, we adopt the class-mixed sampling strategy in DACS to augment the target domain. The best results are denoted in bold. † is the performance reported in the DACS [22]. * is the peak performance model publicly provided by the author of DACS [22].

Method	Road	SW	Build	Wall	Fence	Pole	TL	TS	Veg	Terrain	Sky	Person	Rider	Car	Truck	Bus	Train	MC	Bike	mIoU
ADVENT [24]	89.4	33.1	81.0	26.6	26.8	27.2	33.5	24.7	83.9	36.7	78.8	58.7	30.5	84.8	38.5	44.5	1.7	31.6	32.4	45.5
FDA [28]	92.5	53.3	82.4	26.5	27.6	36.4	40.6	38.9	82.3	39.8	78.0	62.6	34.4	84.9	34.1	53.1	16.9	27.7	46.4	50.5
IAST [9]	**93.8**	**57.8**	85.1	**39.5**	26.7	26.2	43.1	34.7	84.9	32.9	88.0	62.6	29.0	87.3	39.2	49.6	**23.2**	34.7	39.6	51.5
DACS [22]†	89.90	39.66	87.87	30.71	**39.52**	38.52	**46.43**	52.79	**87.98**	43.96	**88.76**	**67.20**	**35.78**	84.45	45.73	50.19	0.00	27.25	33.96	52.14
DACS [22]*	93.25	50.20	87.21	36.75	34.80	38.83	39.80	48.68	87.06	44.06	88.76	65.19	34.38	89.25	**51.64**	52.71	0.00	28.59	48.42	53.66
Ours (DACS+UCT)	93.03	55.92	**87.91**	38.19	38.76	**40.44**	42.14	**54.50**	87.53	**46.67**	87.77	66.26	33.67	**90.18**	47.54	**54.15**	0.00	**41.24**	**53.34**	**55.75**

Setup. We adopt the GTA5 dataset as the source domain, and the Cityscapes dataset as the target domain. The label space of source domain is composed of *road*, *sidewalk*, *building*, *wall*, *fence*, *pole*, *traffic light*, *traffic sign*, *terrain*, *vegetation*, *sky*, *moving objects*. The *moving objects* class in the source domain is further divided into 8 classes, including *person*, *rider*, *car*, *truck*, *bus*, *train*, *motorcycle* and *bicycle* in the target domain.

Comparison with the SOTA. In Table S1, we show the quantitative comparison between our proposed method, the non-adapted baseline “source” and other SOTA self-training based method IAST [9]. Same as the “source” baseline in the Table 2, Table 3 and Table 4 of the main paper, the non-adapted baseline “source” in Table S1 is trained in the supervised way on the labeled source domain and the few-shot labeled target domain. It is shown that both of the adaptation-based methods, IAST and our proposed method, perform better than the non-adapted baseline method, 48.08%, 58.87% *vs.* 32.13%. Moreover, our proposed method outperforms the IAST method by a large margin, 58.87% *vs.* 48.08%. It proves the effectiveness of our proposed method when dealing with the inconsistent taxonomy involving more classes.

S5.2 UDA: Consistent Taxonomy

In Table 1 of the main paper, we show the comparison between our suggested mixed-sampling and contrastive learning based scheme and other SOTA methods under traditional UDA setting, SYNTHIA→Cityscapes. It is shown that our suggested mixed-sampling and contrastive learning based scheme outperforms other SOTA methods under traditional UDA setting. Here we provide

Table S3: More baselines, under Table 2 setting in the main paper.

Method	FSS [30]	ASS [26]	BUDA [1]	DDM [3]	MME [18]	OpenMatch [17]	Ours (BMS)	Ours(All)
mIoU	23.76	24.10	33.63	34.27	31.58	31.23	**43.44**	**49.72**

additional quantitative experimental results under the traidtional UDA setting, GTA5→Cityscapes, to further prove the effectiveness of our suggested mixed-sampling and contrastive learning based scheme for traditional UDA problem.

Setup. We adopt the GTA5 dataset as the source domain, and the Cityscapes dataset as the target domain. The source domain and the target domain share the same label space, where there are 19 classes in total: *road*, *sidewalk*, *building*, *wall*, *fence*, *pole*, *traffic light*, *traffic sign*, *vegetation*, *terrain*, *sky*, *person*, *rider*, *car*, *truck*, *bus*, *train*, *motorcycle* and *bicycle*. We strictly follow the traditional UDA setting, and the target domain is completely unlabelled.

Comparison with the SOTA. In Table S2, we report the quantitative experimental results of our suggested mixed-sampling and contrastive learning based scheme and other SOTA methods under the traditional UDA setting. It is shown that our suggested mixed-sampling and contrastive leaning based scheme outperforms current SOTA methods under the traditional UDA setting, 55.75% *vs.* 53.66%. It further verifies the validity of our suggested mixed-sampling and contrastive learning based scheme for traditional UDA problem.

S5.3 Comparisons to Other Contrastive Learning based Methods

Different from the image-wise [25], semantic distribution-wise [8] and categorical dictionary-wise [5] contrasitve learning, our CT module is the pixel-wise contrastive learning, and is further rectified by the uncertainty prediction, *i.e.,* , UCT. Among [5,8,25], [25] is for image classification, while [5,8] are for semantic segmentation. Compared to [5,8] under Table 1 setting of the main paper, our method outperforms [5,8], 51.45% *vs.* 50.2%, 46.0%, proving the advantage of our uncertainty rectified pixel-wise contrastive learning.

S5.4 Comparisons to Other DA Problems Baselines

Though some works, *e.g.,* , [17, 18], explored open-set or few-shot/semi- DA problems, most of the methods were designed for image classification, while only few of them attempt semantic segmentation [1, 3, 26, 30]. Thus, we incorporate the popular and effective fine-tuning [46] and pseudo-label [26] based few-shot training strategy into the SOTA DA segmentation methods as baselines in Table 2-4 of the main paper. To further prove the effectiveness of our proposed approach for TACS problem, we compare to more baselines in Table S3. [1] is a zero-shot/ class-incremental DA method. [3] is the semi-supervised DA method. The dual-level domain mixing in [3] is realized by 1) randomly copying rectangular region from the source image to the well-labeled target image and 2) concatenating them into one mini-batch. DDM [3] suffers from the saturation of the mixed sample though the color jitter and Gaussian blur are applied, due to only few-shot target samples are labeled in TACS setting. Instead, our BMS copies the class-specific regions from both the source and labeled target image to the abundant unlabeled target image, preventing the saturation (43.44% *vs.* 34.27% in Table S3). The semantic-level adaptation (SA) module in [26] is removed under TACS, because the SA module requires the consistent taxonomy between the source classes and the labeled target classes, while only the inconsistent taxonomy target classes are few-shot labeled under TACS. [18] and [17] are proposed for image classification, thus we modify them for segmentation by, 1) implementing [18] based on ADVENT [24], 2) since [17] does not touch DA problem, we combine [17] with the DA method FDA [28] by applying the core open-set soft-consistency loss in [17] to the images before and after Fourier transform in [28].

S5.5 Additional Qualitative Results

In Fig. 5 of the main paper, we show the qualitative semantic segmentation results, w/o adaptation and adapted with our proposed method, under the open taxonomy, coarse-to-fine taxonomy and implicitly-overlapping taxonomy setting. Here we further provide more qualitative segmentation results, w/o adaptation, adapted with other method, and adapted with our proposed method, under the aforementioned settings. In Fig. S1, under different inconsistent taxonomy settings, we show the qualitative semantic segmentation results on the target domain, w/o adaptation, adapted with IAST [9], and adapted with our proposed method. It is shown that our proposed method outperforms the non-adaptation baseline and other adaptation-based method IAST [9] qualitatively. It further proves the effectiveness of our proposed method for the TACS problem.

Fig. S1: Qualitative semantic segmentation results on the target domain under different inconsistent taxonomy settings, open taxonomy, coarse-to-fine taxonomy and implicitly-overlapping taxonomy. (a) shows the RGB target domain image. (b) gives the ground truth semantic segmentation map. (c) is the semantic segmentation result without adaptation. (d) is the semantic segmentation result adapted by the IAST [9] method. (e) is the semantic segmentation result adapted by our proposed method. Refer to the red box region for the adaptation results of the inconsistent taxonomy classes. The target domain label space of open taxonomy and coarse-to-fine taxonomy setting both have 19 classes, whose corresponding color in the semantic segmentation map is listed in the top color grid. The target domain label space of the implicitly-overlapping taxonomy setting has 16 classes, whose corresponding color in the semantic segmentation map is listed in the low color grid.

References

1. Bucher, M., Vu, T.H., Cord, M., Pérez, P.: Handling new target classes in semantic segmentation with domain adaptation. arXiv preprint arXiv:2004.01130 (2020) 2, 3, 7
2. Cao, Z., Ma, L., Long, M., Wang, J.: Partial adversarial domain adaptation. In: ECCV (2018) 2
3. Chen, S., Jia, X., He, J., Shi, Y., Liu, J.: Semi-supervised domain adaptation based on dual-level domain mixing for semantic segmentation. In: CVPR (2021) 7
4. Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., Schiele, B.: The cityscapes dataset for semantic urban scene understanding. In: CVPR (2016) 4
5. Huang, J., Guan, D., Xiao, A., Lu, S., Shao, L.: Category contrast for unsupervised domain adaptation in visual tasks. In: CVPR (2022) 7
6. Kundu, J.N., Venkatesh, R.M., Venkat, N., Revanur, A., Babu, R.V.: Class-incremental domain adaptation. In: ECCV (2020) 2, 3
7. Lambert, J., Liu, Z., Sener, O., Hays, J., Koltun, V.: MSeg: A composite dataset for multi-domain semantic segmentation. In: CVPR (2020) 2
8. Li, S., Xie, B., Zang, B., Liu, C.H., Cheng, X., Yang, R., Wang, G.: Semantic distribution-aware contrastive adaptation for semantic segmentation. arXiv preprint arXiv:2105.05013 (2021) 7
9. Mei, K., Zhu, C., Zou, J., Zhang, S.: Instance adaptive self-training for unsupervised domain adaptation. In: ECCV (2020) 4, 6, 8, 9
10. Motiian, S., Jones, Q., Iranmanesh, S.M., Doretto, G.: Few-shot adversarial domain adaptation. In: NeurIPS (2017) 2
11. Ning, M., Lu, D., Wei, D., Bian, C., Yuan, C., Yu, S., Ma, K., Zheng, Y.: Multi-anchor active domain adaptation for semantic segmentation. In: ICCV (2021) 1
12. Panareda Busto, P., Gall, J.: Open set domain adaptation. In: ICCV (2017) 2, 3
13. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: NeurIPS (2019) 4
14. Prabhu, V., Chandrasekaran, A., Saenko, K., Hoffman, J.: Active domain adaptation via clustering uncertainty-weighted embeddings. In: ICCV (2021) 1
15. Richter, S.R., Vineet, V., Roth, S., Koltun, V.: Playing for data: Ground truth from computer games. In: ECCV (2016) 4
16. Ros, G., Sellart, L., Materzynska, J., Vazquez, D., Lopez, A.M.: The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In: CVPR (2016) 4
17. Saito, K., Kim, D., Saenko, K.: Openmatch: Open-set semi-supervised learning with open-set consistency regularization. In: NeurIPS (2021) 7, 8
18. Saito, K., Kim, D., Sclaroff, S., Darrell, T., Saenko, K.: Semi-supervised domain adaptation via minimax entropy. In: ICCV (2019) 7, 8
19. Saito, K., Yamamoto, S., Ushiku, Y., Harada, T.: Open set domain adaptation by backpropagation. In: ECCV (2018) 2, 3
20. Shin, I., Kim, D.J., Cho, J.W., Woo, S., Park, K., Kweon, I.S.: Labor: Labeling only if required for domain adaptive semantic segmentation. In: ICCV (2021) 1
21. Teshima, T., Sato, I., Sugiyama, M.: Few-shot domain adaptation by causal mechanism transfer. In: ICML (2020) 2

22. Tranheden, W., Olsson, V., Pinto, J., Svensson, L.: Dacs: Domain adaptation via cross-domain mixed sampling. In: WACV (2021) 4, 6
23. 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) 2
24. Vu, T.H., Jain, H., Bucher, M., Cord, M., Pérez, P.: Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In: CVPR (2019) 2, 6, 8
25. Wang, R., Wu, Z., Weng, Z., Chen, J., Qi, G.J., Jiang, Y.G.: Cross-domain contrastive learning for unsupervised domain adaptation. IEEE Transactions on Multimedia (2022) 7
26. Wang, Z., Wei, Y., Feris, R., Xiong, J., Hwu, W.M., Huang, T.S., Shi, H.: Alleviating semantic-level shift: A semi-supervised domain adaptation method for semantic segmentation. In: CVPR Workshops (2020) 7, 8
27. Wrenninge, M., Unger, J.: Synscapes: A photorealistic synthetic dataset for street scene parsing. arXiv preprint arXiv:1810.08705 (2018) 4
28. Yang, Y., Soatto, S.: Fda: Fourier domain adaptation for semantic segmentation. In: CVPR (2020) 4, 6, 8
29. You, K., Long, M., Cao, Z., Wang, J., Jordan, M.I.: Universal domain adaptation. In: CVPR (2019) 2, 3
30. Zhang, J., Chen, Z., Huang, J., Lin, L., Zhang, D.: Few-shot structured domain adaptation for virtual-to-real scene parsing. In: ICCV Workshops (2019) 2, 7