Supplementary Material - Learn from the Learnt: Source-Free Active Domain Adaptation via Contrastive Sampling and Visual Persistence

Mengyao Lyu[1,2*], Tianxiang Hao[1,2*], Xinhao Xu[1,2], Hui Chen[1,2✉], Zijia Lin[1], Jungong Han[1,2], and Guiguang Ding[1,2✉]

[1] Tsinghua University, Beijing, China
[2] BNRist, Beijing, China
mengyao.lyu@outlook.com, beyondhtx@gmail.com, {xxh22@mails, huichen@mail}.tsinghua.edu.cn, linzijia07@tsinghua.org.cn, jungonghan77@gmail.com, dinggg@tsinghua.edu.cn

Fig. 1: t-SNE visualization of unlabeled target samples (colored by classes) and actively queried samples (marked by cross) on VisDA-C with 0.1% budget per round.

A Qualitative Analysis

A .1 t-SNE Visualization

To visualize the task model's comprehension of the target data and the selection of informative samples during the iterative process of active learning and domain adaptation, we present the t-SNE [2] plots for the VisDA-C dataset in Fig. 1. For the sake of visual clarity, only a subset of unlabelled data is randomly sampled and depicted in each round.

We first observe that, as the volume of actively selected samples increases, the classification boundaries are progressively revealed, indicating that the task model's understanding of the target data distribution has become more comprehensive. As our

* Equal Contribution ✉ Corresponding Author

CAS is motivated, under the domain shift problem, unlabeled samples we identify as challenging to transfer are proximate to decision boundaries in the target domain, where our active selection often happens. At the same time, our sampling process prioritizes minority classes (e.g., knife), which is contributed to our class-level transferability estimation.

A .2 Queried Target Samples

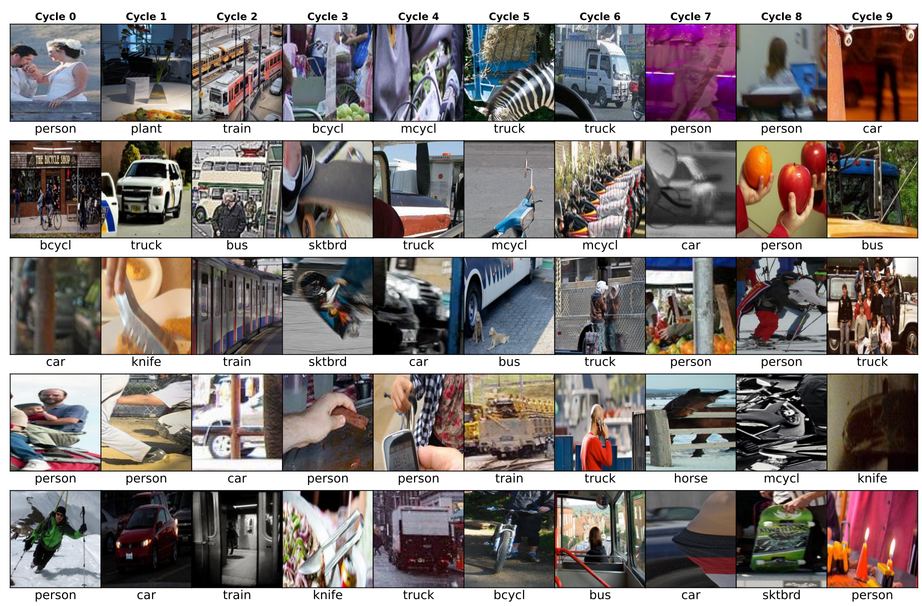

Fig. 2: Top-5 CAS samples of 10 rounds on VisDA-C.

We further validate the efficacy of proposed CAS via visual inspection of the queried target samples. Fig. 2 shows top-5 actively queried samples of each active learning round on VisDA-C. Based on the recognition results from the source-only model, person (19.2%), truck (6.1%) and knife (4.9%) are challenging classes. We can observe that more iconic and representative samples are selected in order to learn authentic features that differ from those of the generated images. On the other hand, for more transferable classes, e.g., car (80.0%) and motorcycle, samples that are truncated, occluded, distant and blurry are queried at first. Those patterns are not present in the source domain but are ubiquitous in real world scenarios, which provide rich information for the adaptation.

We have also noted that in the later stages of iteration, our CAS not only selects samples that are prone to confusion but is also able to identify instances with inaccurate labels. For example, the image where human hands holding apples is labeled as person, or the case where the foreground is a person but the label corresponds to a blurry truck in the background.

Table 1: Ablation study with 1%-labeled VisDA-C and 5%, 10%-labeled Office-31.

u_{cm}	u_{ct}	$\mathcal{L}_{ac}$	$\mathcal{L}_{vpa}$	1%-VisDA-C	5%-O31	10%-O31
				85.1	89.2	93.1
✓				86.7	91.4	94.8
✓	✓			86.9	91.8	95.1
✓	✓	✓		87.0	92.0	95.3
✓	✓		✓	**87.4**	**93.5**	**95.4**

Table 2: Results of applying CAS to SFUDA SOTAs on VisDA-C.

AS	CAS	SHOT	SHOT++	SF(DA)^2	MHPL	LFTL
0	✗	85.5	87.3	<u>88.1</u>	-	-
1	✓	86.4	87.5	<u>88.9</u>	-	87.4
5	✓	91.9	91.8	92.3	91.3	<u>92.8</u>

B Component Verification

B .1 Ablation Study

Take the combination of vannila BvSB [1] sampling strategy and labeled target supervision via CE loss as a baseline, we present the ablation study of the propose LFTL framework in Tab. 1. On VisDA-C and Office-31 datasets with differing budget constraints, consistent improvements can be observed with each component, which validate our motivations. Given the same annotation budget, our CAS strategy prioritizes target samples that remain unrecognizable to the current model and have not been captured in preceding active learning rounds, and meantime it factors out samples with knowledge previously acquired when the hypotheses exhibit increased confidence. In addition to the contrastive margin u_{cm}, the class-level transferability u_{ct} enhances our sampling criterion with a global semantic perspective, which promotes the selection of tail and challenging classes. The actively selected target samples then play the role of anchors to guide the optimization during the adaptation procedure. Without explicit error-prone pseudo-labeling or time-consuming clustering, the $\mathcal{L}_{ac}$ efficiently encourages density around active anchors. When replaced with features with VP delivered via $\mathcal{L}_{VPA}$, the representations of active anchors are features temporally ensembled from the source model to the current, which efficiently promotes a source-like feature distribution in the target domain.

B .2 CAS for SFUDA SOTAs

This section further verifies CAS by applying it to SFUDA methods to facilitate their adaptation. Here we take SOTAs SHOT and SF(DA)^2 for example. Results in Tab. 2 show that CAS can be simply plugged into source-free methods to query informative target samples for a performance boost. Meantime, a better approach is to simultaneously

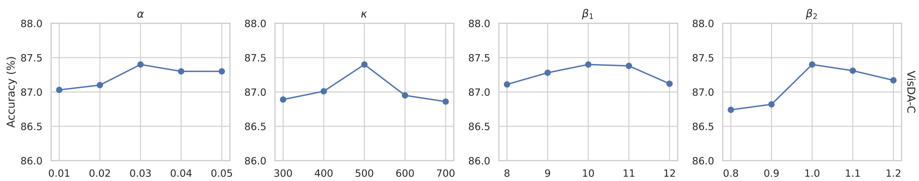

Fig. 3: Parameter sensitivity analysis of α in Eq. 1, κ in Eq. 3, and β_1, β_2 in Eq. 10 on VisDA-C.

consider the risk of forgetting the domain-invariant knowledge during transfer and thereby incorporating our VPA strategy to learn from the learnt.

B .3 Parameter Sensitivity Analysis

To validate the effectiveness and generalization ability of the proposed method, we study the sensitivity of LFTL to α in Eq. 1, κ in Eq. 3, and β_1, β_2 in Eq. 10 on the VisDA-C dataset. We experiment around the optimal values of parameters, perform three trials with a set of seeds and average the results. In Fig. 3, we observe similar bell-shaped curves on all experiments, indicating consistent performance gains benefit from our proposed methods and demonstrating robustness to parameter choices.

C Additional Implementation Details

During the sample-and-adaptation interplay, the SGD optimizer with a momentum of 0.9 and a weight decay of $1e{-}3$ is applied. The learning rate is set as $1e{-}3$ for VisDA-C and $1e{-}2$ for Office, scheduled by $\eta = \eta_0(1 + 10p)^{-0.75}$ where p increases from 0 to 1 during optimization. We set batch size as 64. For all datasets, the CAS coefficient $\alpha = 0.03$, β_1, β_2 in model adaptation are set to 10.0 and 0.9 to balance the magnitude of each loss terms. We set $\kappa = 500$ for the large-scale VisDA-C and 100 for smaller Office datasets.

References

1. Lewis, D.D., Gale, W.A.: A sequential algorithm for training text classifiers. In: Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval. pp. 3–12. Springer (1994)
2. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)