Cross-Domain Adaptive Clustering for Semi-Supervised Domain Adaptation
Supplementary Document

Jichang Li[1] Guanbin Li[2*] Yemin Shi[3] Yizhou Yu[1*]
[1]The University of Hong Kong [2]Sun Yat-sen University [3]Deepwise AI Lab
csjcli@connect.hku.hk, liguanbin@mail.sysu.edu.cn, shiyemin@deepwise.com, yizhouy@acm.org

We present additional analysis of our proposed approach, Cross-domain Adaptive Clustering (CDAC), in this supplementary document.

1. Additional Analysis

Additional performance comparisons on the DomainNet benchmark: We show additional comparisons with a varying number of labeled target domain samples of each category, *i.e.*, 5-shot and 10-shot per class, on the *DomainNet* benchmark using Resnet34 as the backbone network in Table 1. In comparison to the existing state-of-the-art SSDA approaches, the proposed method achieves better classification performance on *DomainNet* in all adaptation scenarios. Specifically, our CDAC method outperforms the previous best results by 3.3% and 3.0% on average under the 5-shot and 10-shot settings respectively.

Effectiveness of Adversarial Adaptive Clustering: Inspired by [2, 3], Cluster Core Distance (CCD) is introduced to evaluate the effectiveness of the adversarial adaptive clustering loss, which measures the distance between the two source and target domain feature clusters within the same class. In details, the CCDs can be defined as $\{d^e_1, d^e_2, \dots, d^e_k, \dots, d^e_K\}$, where d^e_k denotes the Euclidean distance [1] for class k in the e-th epoch during model training and K represents the number of classes in a dataset. To be fair, the CCD d^e_k at each epoch e is normalized by the initial CCD d^0_k, which is calculated using the initial model parameterized by pre-trained weights on ImageNet without any fine-tuning. In general, the more aligned cross-domain feature clusters are, the smaller CCDs are.

We use 2000 samples (1000 from source domain and 1000 from target domain, each includes 50 samples per class from 20 representative classes) in this validation experiment on *DomainNet* in the adaptation scenario, "R→S", under the 3-shot setting using Resnet34 as the backbone. We compare our CDAC model with "Pre-trained", "S+T" and "CDAC w/o PL" models. Specifically, "Pre-trained" means that the model is parameterized using pre-trained weights on ImageNet without any further training, while the "S+T" model is trained with labeled samples in the source and target domains only. Also, "CDAC w/o PL" denotes a degraded version of our complete CDAC model, and is trained with the standard cross-entropy loss and the proposed adversarial adaptive clustering loss without using any pseudo labels. We show the final CCDs of the above mentioned 20 representative classes in Figure 1, and it can be observed that in every class, the CCD of our complete CDAC model achieves the smallest value. These results further verify that our proposed approach can effectively perform cross-domain cluster-wise feature alignment and help improve the classification performance of SSDA models.

*Corresponding Authors.

References

[1] Jon Dattorro. *Convex optimization & Euclidean distance geometry*. Lulu. com, 2010. 1

[2] Yawei Luo, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2507–2516, 2019. 1

[3] Chaofan Tao, Fengmao Lv, Lixin Duan, and Min Wu. Minimax entropy network: Learning category-invariant features for domain adaptation. *arXiv preprint arXiv:1904.09601*, 2019. 1

Table 1. Performance comparisons on *DomainNet* under the 5-shot and 10-shot settings using Resnet34 as backbone.

Net	Method	R→C	R→P	P→C	C→S	S→P	R→S	P→R	MEAN
5-shot									
Resnet34	S+T	64.5	63.1	64.2	59.2	60.4	56.2	75.7	63.3
	DANN	63.7	62.9	60.5	55.0	59.5	55.8	72.6	61.4
	ENT	77.1	71.0	75.7	61.9	66.2	64.6	81.1	71.1
	MME	75.5	70.4	74.0	65.0	68.2	65.5	79.9	71.2
	APE	77.7	73.0	76.9	67.0	71.4	68.8	80.5	73.6
	CDAC	**80.8**	**75.3**	**79.9**	**72.1**	**74.7**	**72.9**	**83.2**	**76.9**
10-shot									
Resnet34	S+T	68.5	66.4	69.2	64.8	64.2	60.7	77.3	67.3
	DANN	70.0	64.5	64.0	56.9	60.7	60.5	75.9	64.6
	ENT	79.0	72.9	78.0	68.9	68.4	68.1	82.6	74.0
	MME	77.1	71.9	76.3	67.0	69.7	67.8	81.2	73.0
	APE	79.8	75.1	78.9	70.5	73.6	70.8	82.9	75.9
	CDAC	**83.1**	**77.2**	**81.7**	**74.3**	**76.3**	**74.6**	**84.7**	**78.9**

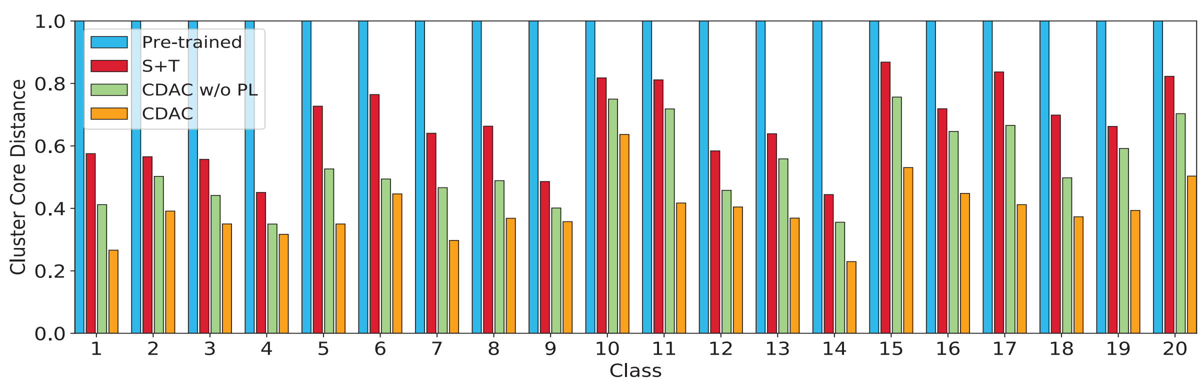

Figure 1. Comparison of final cluster core distances (CCDs) of 20 representative classes. Class 1-20 denote “see_saw”, “speedboat”, “sheep”, “leaf”, “raccoon”, “feather”, “laptop”, “dog”, “umbrella”, “grapes”, “streetlight”, “foot”, “butterfly”, “axe”, “eyeglasses”, “goatee”, “drums”, “helmet”, “asparagus” and “penguin”. For each class, we show four results obtained from different approaches to indicate different approaches have different abilities to make target domain clusters and their corresponding source domain clusters closer. Apparently, our CDAC approach produces more discriminative features to help align cluster-wise feature distributions across domains.