CDS: Cross-Domain Self-supervised Pre-training Supplementary Material

Donghyun Kim[1], Kuniaki Saito[1], Tae-Hyun Oh[2], Bryan A. Plummer[1], Stan Sclaroff[1], Kate Saenko[1,3]
[1]Boston University, [2]POSTECH, [3]MIT-IBM Watson AI Lab
{donhk, keisaito, bplum, sclaroff, saenko}@bu.edu, taehyun@postech.ac.kr

In this supplementary material, we provide additional details and results, which we cannot show in the main paper due to the limited space.

A. Dataset Statistics

Table A shows the overall statistics of the datasets including the number of domains, images, and classes. 52% of the classes in Office-Home overlap with ImageNet classes. However, there are similar classes across datasets (e.g., table and desk). 59 high-level bird classes are in ImageNet. There are only 10 high-level bird classes in ImageNet that overlap with fine-grained bird classes in CUB-200 by matching high-level class names.

Dataset Statistics				
CUB [6, 7]				
Domain	Real (R)	Painting (P)		
# total images	10,788	3,047		
# classes	200	200		
Office-Home [5]				
Domain	Art(Ar)	Clipart (Cl)	Product (Pr)	Real (Rw)
# images	2,427	4,365	4,439	4,357
# classes	65	65	65	65
Office [3]				
Domain	Amazon (A)	Dslr (D)	Webcam (W)	
# images	2,817	498	795	
# classes	31	31	31	

Table A: Dataset statistics of the CUB, Office-Home, and Office datasets used in our experiments.

Pre-train	CUB: Target Acc (%) on 1-shot/3-shots							
	Real→Painting				Painting→Real			
	P@1	P@5	P@15	AVG	P@1	P@5	P@15	AVG
ImageNet	21.6	16.5	14.4	17.5	23.6	21.0	17.9	20.8
ID	18.3	12.9	10.7	14.0	26.4	23.1	19.0	22.8
SimCLR	11.8	10.0	9.1	10.3	14.9	13.2	11.0	13.0
SimCLR+DC	11.7	10.0	9.1	10.3	15.2	13.3	11.0	13.2
In-domain ID	20.5	15.7	13.1	16.4	25.2	22.1	18.5	21.9
CDS	**38.8**	**34.3**	**33.6**	**35.6**	**43.1**	**40.7**	**36.9**	**40.2**

Table B: Detailed results of Table 1 on CUB. We report Precision@k (P@k) of different pre-training methods on the unsupervised cross-domain image retrieval task.

B. Cross-domain Image Retrieval

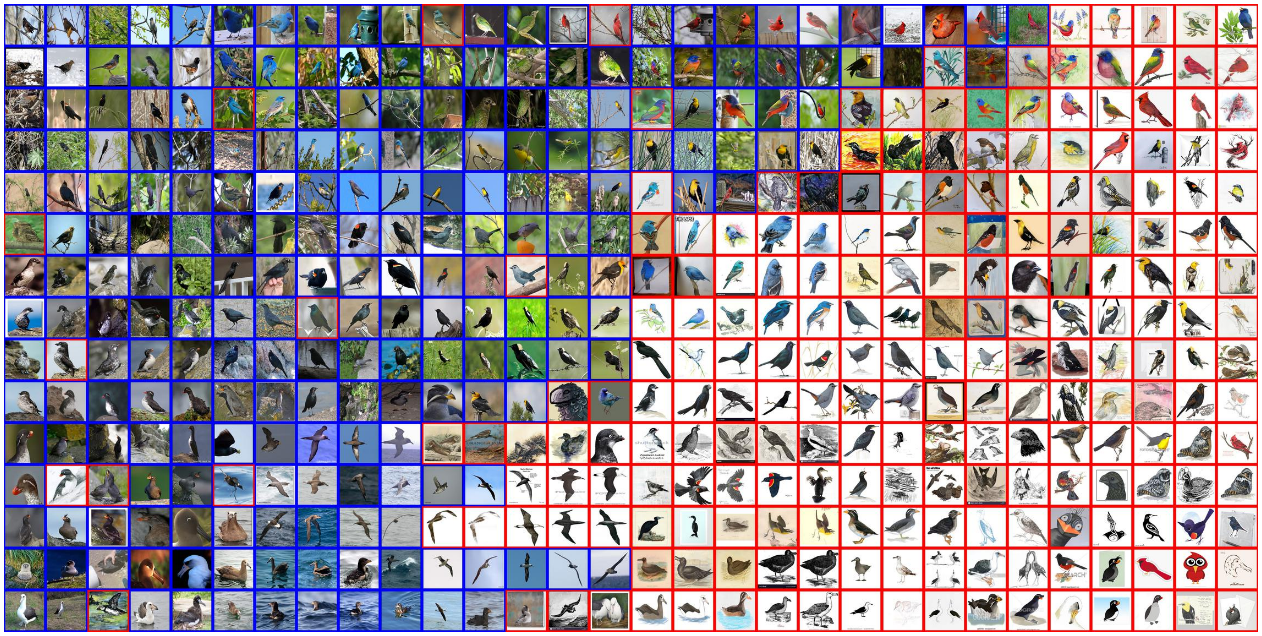
We report the detailed results of Precision@k on CUB in Table B. CDS improves averaged Precision@k by 18.1% and 19.4% on each setting in CUB compared to ImageNet pre-training. We show additional visualization from a ImageNet pre-trained model and ours in Fig. A and B. In Fig. A, based on the features obtained from t-SNE, we show the corresponding images. Red boxes represent the painting domain and blue boxes represent the real images. We can easily see that ImageNet features are biased to its background, so the two domain features are highly separated regardless of their semantic classes. However, CDS focuses more on the object rather than the background, and the images of the same class from the two domains tend to embed nearby each other. This shows that CDS learns more discriminative features that are also domain-invariant. Additional examples of unsupervised cross-domain retrieval on Office-Home can be found in Fig. B. We also show some failure cases of both ImageNet features and ours. These failure examples share very similar colors or shapes between different classes. Learning to discriminate between these similar shapes or colors in images of different classes will be important in future work.

C. Universal Domain Adaptation

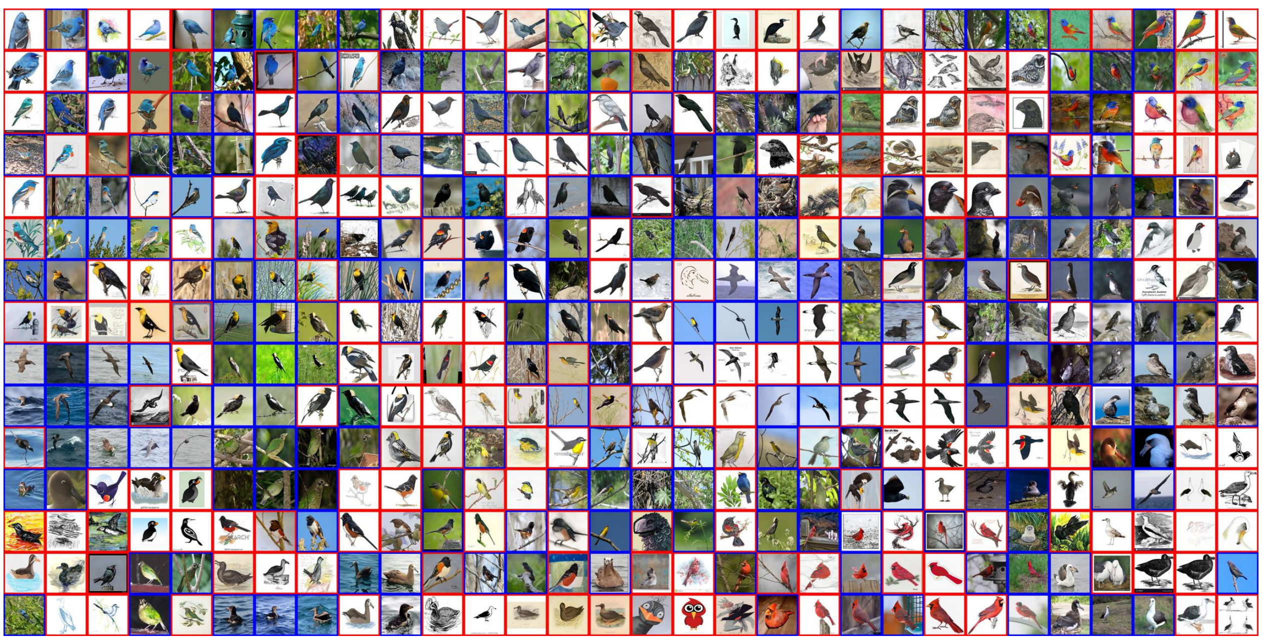
Evaluation metric for open set DA. For open set and open-partial domain adaptation, mean class accuracy is used as an evaluation metric. Mean class accuracy is important as the number of unknown samples can overwhelm the number of known class samples in the target domain. In the mean class accuracy, each class has the contribution as the whole unknown classes, so that the importance of the unknown class can be too small when the number of known classes is large. Fu *et al.* [1] proposed the H-score metric, which is the harmonic mean of the mean class accuracy on known classes and the accuracy on the "unknown" class.

$$h = 2 \cdot \frac{acc_{known} \cdot acc_{unknown}}{acc_{known} + acc_{unknown}} \tag{1}$$

However, as the known class accuracy has the same con-



(a) ImageNet



(b) CDS

Figure A: t-SNE visualization of the ImageNet pre-trained model and ours. Red boxes represent the painting domain and blue boxes represent the real domain. In ImageNet pre-trained features (a), the features of two different domains are highly separated. Therefore, images of the same class across domains are embedded far from each other. However, in (b), CDS produces discriminative features as well as domain-aligned features across domains.

tribution (*i.e.*, importance) as the unknown class in H-score, H-score can put too much weight on the unknown class and may not fully reflect recognition performance on the known classes (*e.g.*, when the number of the unknown classes is too small or the number of known classes is too large).

Label efficiency comparison with SSL. Fig. C compares the results of the baselines to our approach with different fractions of source labels using DANCE [4] on the

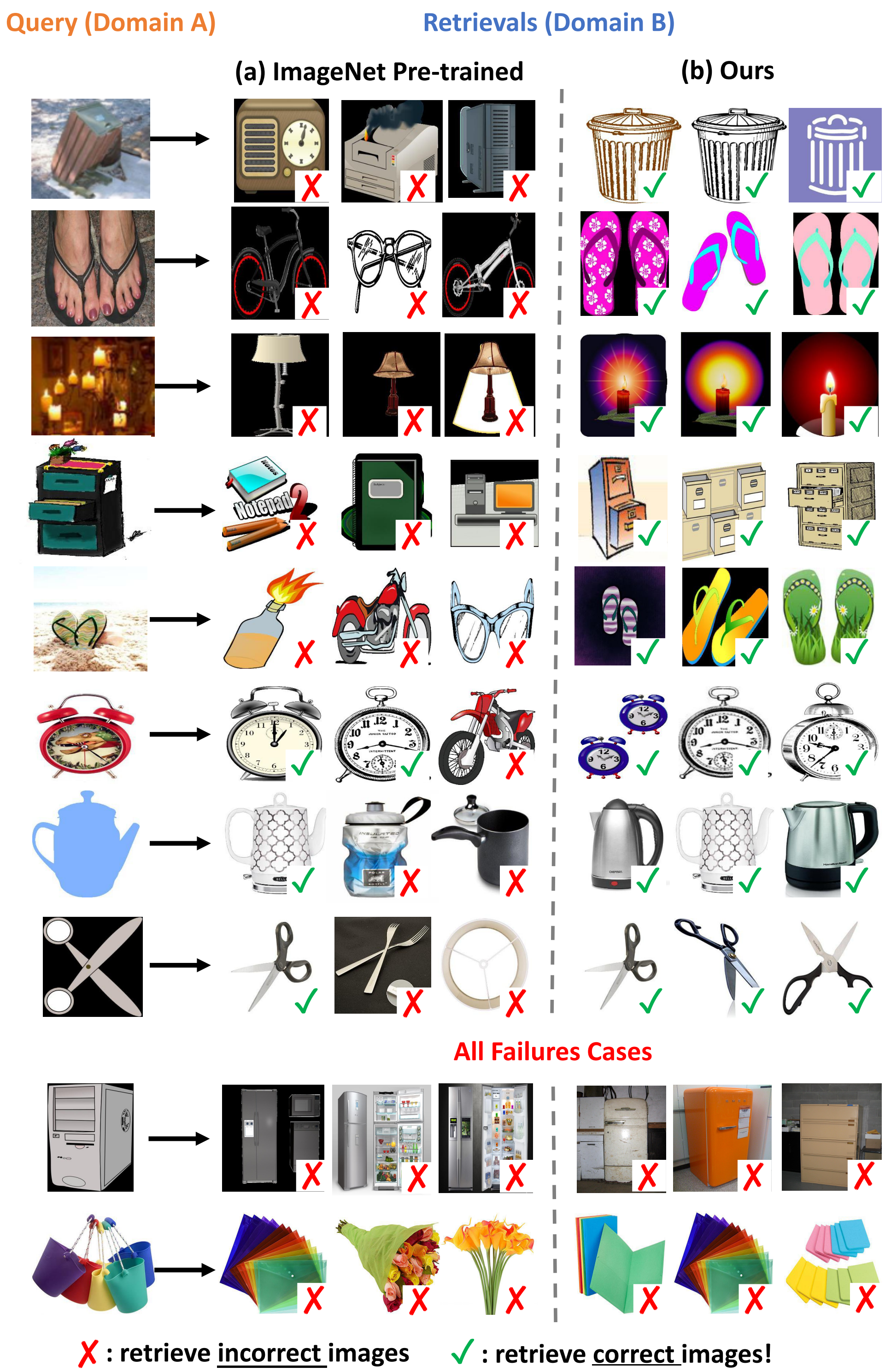


Figure B: Retrieval of the cross-domain neighbors using (a) standard ImageNet pre-trained features and (b) CDS on Office-Home. Similarly, CDS learns better semantic similarity between domains and retrieves correct class images compared to (a) ImageNet pre-trained weights. The last two rows show the examples of failures of both methods.

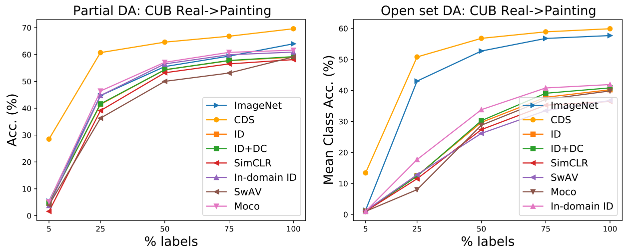
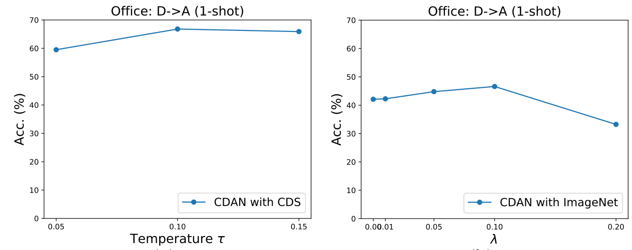


Figure C: Comparison with SSL baselines using DANCE [4] with different fractions of source labels on partial and open set DA.



(a) (b)

Figure D: Sensitivity analysis on the temperature parameter τ and λ on the Office D$\rightarrow$A source 1-shot setting using CDAN.

Real$\rightarrow$Painting setting in CUB under the open set and partial DA settings. We randomly select a subset of source examples as labeled and report the averaged accuracy on three random splits. CDS greatly improves performance when there are limited source labels and consistently outperforms the baselines. In addition, CDS continues to improve with additional labels, while the SSL baselines obtain similar or lower accuracy than ImageNet.

Detailed comparison with SSL baselines on CUB and Office-Home on universal domain adaptation. We report the detailed results of Table 3 in the main paper. Tables C and D show average accuracy of three runs on each DA setting and average standard deviation in the CUB and Office-Home.

Additional results on partial and open set DA with the different number of source and target private classes. Table E shows that results using DANCE with different number of source or target private classes on the Real$\rightarrow$Painting setting in CUB under partial and open set DA. CDS consistently improves upon ImageNet pre-training.

Detailed results on open-partial DA. We report the detailed results of Table 5 in the main paper. Tables F and G show H-score and mean class accuracy on all settings in Office-Home and Office datasets. We also show the averaged standard deviation of three runs on all settings.

D. Few-shot Domain Adaptation

Additional comparison with [2]. We compare our approach with Menapace *et al.* [2] in Table H. Menapace *et al.* [2] proposed an unsupervised clustering method for multiple unlabeled source domains (UCDS) for a domain generalization task, which is different from our task. We explore their method on our few-shot domain adaptation task. The results show that UCDS cannot outperform the ImageNet pre-trained weights which we consider a baseline that we improve upon. In their paper, they train a model from scratch (*i.e.*, random initialization) and use sobel-filtered images (2 channel input) for data augmentation. Thus, their approach cannot take full benefit of the first stage pre-training on ImageNet. The fact that this kind of data augmentation removes all color information could also be harmful where this information is necessary, such as fine-grained classification on CUB as we observe from SimCLR in Table B. We use their released code with minor modifications. Specifically, we remove the sobel-filtering to match the same input channel as the ImageNet pre-trained weights, initialize the model with the ImageNet pre-trained weights, and use UCDS to pre-train a model on the setting of D$\rightarrow$A in Office. With the pre-trained weights with UCDS, we finetune a model with 1-shot source labels in the same way of Table 7 using CDAN. In Table H, we observe that the method UCDS hurts the performance of the ImageNet weights by -15.5%, while our method improves ImageNet weights by 20.2%.

Sensitivity analysis. We use the same hyper-parameters in CDS tuned on one setting and use it for other settings (*e.g.*, tuned on D$\rightarrow$A source 1-shot setting in Office). Figure D-(a) shows the sensitivity analysis of the temperature parameter in CDS. We set the temperature $\tau = 0.1$ for all experiments. We observe that the sensitivity of η is very small and set $\eta = 0.5$ following [8].

For few-shot domain adaptation, we apply a learning objective function of a domain adaptation method ($\mathcal{L}_{DA}$) (*e.g.*, CDAN, MME, DANN) with labeled source and unlabeled target examples. For the unlabeled source data, we also apply entropy minimization loss ($\mathcal{L}_{ent}$). The overall objective is as follows:

$$\mathcal{L} = \mathcal{L}_{DA}(\mathcal{D}_{sl}, \mathcal{D}_{tu}) + \lambda \mathcal{L}_{ent}(\mathcal{D}_{su}) \tag{2}$$

Figure D-(b) shows the sensitivity of λ on entropy minimization for CDAN using the ImageNet pre-trained network. For fair comparison, we apply the same $\lambda = 0.1$ for CDAN to the Imagenet pre-trained network and all SSL baselines.

Domain confusion loss analysis. In Fig. E, we report the confusion loss from the domain classifier used in the CDAN method according to training iterations on the Office D$\rightarrow$A source 1-shot setting. The confusion loss indicates how the source and target features are aligned with each other. The loss decreases at the early stages of training by the discriminator but increases later on by encouraging the feature extractor to confuse the discriminator. CDS obtains a higher

Pre-train	Closed set			Open set (Mean Class Acc.)			Partial set		
	Real→Painting	Painting→Real	AVG	Real→Painting	Painting →Real	AVG	Real→Painting	Painting →Real	AVG
ImageNet	60.2	48.7	54.5	57.7	51.4	54.6	64.0	52.2	58.1
ID	63.1	48.1	55.6	40.2	40.8	40.5	59.0	50.4	54.7
ID+DC	63.1	48.8	56.0	40.8	35.3	38.1	59.2	50.4	54.8
SimCLR	51.3	46.9	49.1	36.4	30.0	33.2	58.1	47.3	52.7
MoCo	60.9	43.9	52.4	36.8	27.7	32.2	59.1	46.0	52.5
Swav	62.1	48.8	55.4	39.8	35.8	35.8	61.6	50.7	56.1
In-domain ID	64.2	49.1	56.6	41.9	37.0	39.5	60.9	51.5	56.2
CDS	**64.3**±0.2	**53.6**±0.3	**59.0**	**59.9**±0.6	**52.1**±0.2	**55.9**	**69.6**±0.5	**61.3**±0.2	**65.4**

Table C: Detailed results of Table 3 using DANCE on CUB. We report averaged accuracy on three trials. For open-set, we report mean class accuracy. We also show the standard deviation over three runs. We observe that the gain of CDS on CUB is larger than that of Office-Home. This is because CUB has many novel categories and bigger category shift from ImageNet.

Pre-train	Office-Home: Target Acc. (%) with few target labels												
	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	AVG
(a) Closed set													
ImageNet	54.3	**75.9**	**78.4**	64.8	72.1	73.4	63.2	53.0	79.4	73.0	58.2	**82.9**	69.1
ID	51.1	71.9	76.1	60.3	69.2	72.0	62.6	48.2	76.3	69.0	59.0	80.5	66.3
ID+DC	50.8	72.1	76.0	60.2	69.4	71.5	61.1	47.9	76.3	69.1	58.4	80.4	66.1
SimCLR	52.0	73.2	76.0	60.1	69.1	70.3	64.2	47.9	76.6	70.2	58.2	80.8	66.6
MoCo	52.2	72.8	75.8	59.6	68.6	69.8	60.8	47.6	75.8	70.2	57.4	80.6	65.9
Swav	52.6	73.3	76.2	61.4	69.9	70.9	66.0	49.5	77.2	70.7	59.3	81.5	67.4
In-domain ID	51.5	72.2	76.1	59.3	69.7	71.3	62.1	48.8	76.5	69.4	59.4	80.5	66.4
CDS	**55.9**	74.6	77.7	**65.5**	**73.3**	**75.0**	**67.8**	**54.5**	**79.5**	**73.7**	**59.3**	82.4	**69.9** (±0.3)
(b) Open set (Mean Class Accuracy)													
ImageNet	64.1	84.1	88.3	**76.7**	80.7	**84.9**	**77.6**	62.7	85.4	80.8	65.1	**87.1**	78.1
ID	56.2	76.8	89.6	67.3	73.3	82.3	60.1	48.6	84.6	74.5	56.7	81.6	71.0
ID+DC	55.5	76.6	89.7	67.6	73.3	82.2	60.9	48.2	85.0	74.7	56.3	81.9	71.0
SimCLR	57.4	79.5	89.9	67.3	74.6	77.2	65.7	50.1	85.7	76.7	56.6	82.1	71.9
MoCo	56.7	78.6	89.9	65.6	73.0	81.4	63.1	48.6	85.8	76.1	54.6	82.4	71.3
Swav	56.2	78.0	90.1	69.4	75.3	82.3	64.6	49.5	85.8	76.7	57.9	81.6	72.3
In-domain ID	65.7	79.1	80.6	72.3	76.9	78.6	61.0	49.5	85.5	75.4	56.4	82.4	72.0
CDS	**66.8**	**84.4**	**90.1**	74.4	**81.0**	84.4	77.0	**65.5**	**87.3**	**81.0**	**66.8**	86.1	**78.7** (±0.3)
(c) Partial													
ImageNet	53.6	**73.2**	**84.9**	70.8	**67.3**	**82.6**	70.0	50.9	**84.8**	77.0	55.9	**81.8**	**71.1**
ID	50.6	65.9	82.7	67.2	62.9	79.3	62.6	47.3	80.9	74.6	54.3	79.5	67.3
ID+DC	50.4	65.9	83.0	67.3	62.9	78.7	61.1	46.7	80.7	74.6	53.8	79.6	67.0
SimCLR	53.5	66.3	84.1	67.6	63.4	78.1	64.2	46.5	82.3	74.3	56.7	80.0	68.1
MoCo	52.4	65.2	83.6	65.6	61.5	76.8	60.8	45.3	81.9	73.7	54.0	78.6	68.6
Swav	53.6	67.9	84.1	68.3	64.3	80.0	66.0	49.0	82.3	75.8	55.6	80.8	69.0
In-domain ID	50.5	65.8	82.8	66.8	62.6	77.7	61.8	45.4	80.3	74.1	55.5	78.8	66.8
CDS	**54.3**	67.4	78.7	**73.8**	63.7	79.9	**71.1**	**53.2**	78.2	**78.2**	**58.0**	79.9	69.7 (±0.5)

Table D: Detailed results of Table 3 using DANCE [4] on Office-Home. We report averaged accuracy on three trials. For open-set, we report mean class accuracy. We also show the averaged standard deviation on all settings in Office-Home over three runs.

confusion loss than ImageNet pre-trained weights, which is further evidence that shows our features are more domain-invariant.

Comparison with MMD. We try to In-domain ID with the maximum mean discrepancy for domain alignment, which obtains 62.1% (6.4% lower than CDS) from a kNN classifier in Table 7.

Detailed results on CUB, Office, and Office-Home on few-shot domain adaptation with few source labels. We report the detailed results of Table 6 in the main paper. Tables I, K, and J show average accuracy of three runs on each DA setting and averaged standard deviations in the CUB, Office-Home, and Office datasets.

E. Additional Evaluation on Pre-trained Models

We additionally evaluate pre-traind models with the LogME metric [9], which can assess the transferability of pre-trained models for target tasks. We measure LogME of ours and the Image-Net pre-trained model. We averaged the value of LogME on all DA scenarios in Office-Home. CDS obtains 1.10 and the Image-Net pre-trained model obtains 1.04.

CUB: Real→Painting				
Pre-train	Partial DA ($|C|/|\bar{C}_s|/|\bar{C}_t|$)			
	120 / 80 / 0	140 / 60 /0	160 / 40 / 0	180 /20 / 0
ImageNet	60.7	58.5	57.1	57.8
CDS	**65.9**	**63.1**	**61.5**	**62.9**
Pre-train	Open set DA ($|C|/|\bar{C}_s|/|\bar{C}_t|$)			
	100 / 0 / 20	100 / 0 / 40	100 / 0 / 60	100 / 0 / 80
(a) H-score				
ImageNet	28.2	25.9	31.6	29.8
CDS	**41.2**	**44.1**	**48.8**	**42.1**
(b) Mean Class Acc.				
ImageNet	59.0	58.6	58.2	57.7
CDS	**60.5**	**60.2**	**58.8**	**58.7**

Table E: Additional results with the different number of source private and target private classes on partial and open set DA using DANCE. We report overall accuracy for partial DA. We report mean class accuracy and H-score for open set DA. $|C|/|\bar{C}_s|/|\bar{C}_t|$ represents the number of shared classes, source private classes, and target private classes respectively.

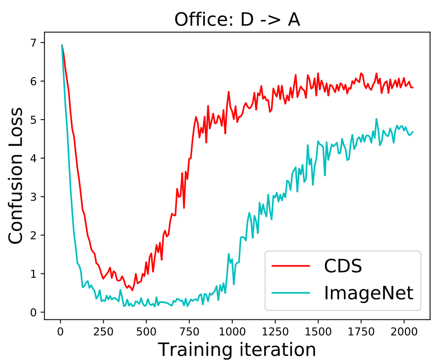


Figure E: Confusion loss (measured with the domain classifier) when using the pre-trained weights obtained by ImageNet and CDS on the Office D→A source 1-shot setting. High confusion loss represents the domain-invariant features between the source and target domains. CDS obtains more domain-invariant features.

F. Memory Bank

Memory Bank is memory-efficient and takes only up to 0.02 GB in our experiments. The size of the memory bank can be further reduced with noise-contrastive estimation as done in [8] for a large dataset.

References

[1] Bo Fu, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Learning to detect open classes for universal domain adaptation. In *European Conference on Computer Vision*, pages 567–583. Springer, 2020. 1

[2] Willi Menapace, Stéphane Lathuilière, and Elisa Ricci. Learning to cluster under domain shift. *arXiv preprint arXiv:2008.04646*, 2020. 4, 7

[3] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *Proceedings of the European Conference on Computer Vision*, pages 213–226. Springer, 2010. 1

[4] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, and Kate Saenko. Universal domain adaptation through self supervision. *arXiv preprint arXiv:2002.07953*, 2020. 2, 4, 5

[5] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5018–5027, 2017. 1

[6] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011. 1

[7] Sinan Wang, Xinyang Chen, Yunbo Wang, Mingsheng Long, and Jianmin Wang. Progressive adversarial networks for fine-grained domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9213–9222, 2020. 1

[8] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3733–3742, 2018. 4, 6

[9] Kaichao You, Yong Liu, Jianmin Wang, and Mingsheng Long. Logme: Practical assessment of pre-trained models for transfer learning. In *International Conference on Machine Learning*, pages 12133–12143. PMLR, 2021. 5

Method	Office-Home												
	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	AVG
						H-score							
SO	44.7	48.0	50.1	46.6	46.9	49.0	47.5	43.2	50.2	48.5	44.8	48.4	47.3
DANN	42.4	48.0	48.9	45.5	46.6	48.4	45.8	42.6	48.7	47.6	42.7	47.4	46.2
UAN	51.6	51.7	54.3	61.7	57.6	61.9	50.4	47.6	61.5	62.9	52.6	65.2	56.6
CMU	56.0	56.9	59.2	67.0	64.3	67.8	54.7	51.1	66.4	68.2	57.9	69.7	61.6
DANCE	48.1	43.5	45.8	55.9	37.4	46.9	60.9	55.4	51.4	40.7	51.1	52.9	49.2
DANCE+CDS	**67.9**	**75.8**	**77.2**	**72.0**	**69.6**	**71.6**	**74.2**	**64.5**	**75.8**	**67.7**	**63.3**	**70.4**	**70.8** (±0.4)
						Mean Class Accuracy							
SO	44.7	48.0	50.1	46.6	46.9	49.0	47.5	43.2	50.2	48.5	44.8	48.4	47.3
DANN	42.4	48.0	48.9	45.5	46.5	48.4	45.8	42.6	48.7	47.6	42.7	47.4	46.2
UAN	51.6	51.7	54.3	61.7	57.6	61.9	50.4	47.6	61.5	62.9	52.6	65.2	56.6
CMU	56.0	56.9	59.2	67.0	64.3	67.8	54.7	51.1	66.4	68.2.	57.9	69.7	61.6
DANCE	64.1	84.3	91.2	**84.3**	78.3	**89.4**	**83.4**	63.6	91.4	83.3	63.9	86.9	80.4
DANCE+CDS	**67.9**	**86.9**	**94.2**	76.5	**78.6**	89.0	80.8	**65.0**	**92.2**	**83.6**	**68.5**	**88.4**	**81.0** (±0.3)

Table F: Detailed results of Table 5 on Office-Home under the open-partial setting. DANCE with CDS slightly improves mean class accuracy on average and significantly improves the H-score on the open-partial settings compared to DANCE with ImageNet pre-training.

Method	Office (H-score / Mean Class Acc.)						
	A→D	A→W	D→A	D→W	W→A	W→D	AVG
SO	49.8 / 80.5	47.9 / 75.9	48.5 / 78.8	54.9 / 89.6	49.0 / 81.4	55.6 / 90.9	50.9 / 82.9
DANN	50.2 / 82.7	48.8 / 80.7	47.7 / 74.8	52.7 / 80.9	49.3 / 83.5	54.9 / 88.1	50.6 / 81.8
UAN	59.7 / 86.5	58.6 / 85.6	60.1 / 85.5	70.6 / 94.8	60.3 / 85.1	71.4 / 98.0	63.5 / 89.2
CMU	68.1 / 89.1	67.3 / 86.9	71.4 / 88.4	79.3 / 95.7	72.2 / 88.6	80.4 / 98.0	73.1 / 91.1
DANCE	78.5 / **91.0**	70.3 / **92.1**	79.3 / **91.9**	90.2 / **97.8**	74.0 / **91.3**	89.6 / **98.0**	80.3 / **93.7**
DANCE+CDS	**83.2** / 84.4	**86.1** / 86.8	**90.6** / 91.5	**90.6** / 96.3	**89.6** / 91.0	**83.9** / 97.4	**87.3** (±0.8) / 91.2 (±0.6)

Table G: Detailed results of Table 5 on Office under the open-partial settings. DANCE with CDS obtains slightly lower mean class accuracy but significantly improves the H-score on the open-partial settings.

Pre-train	Target Acc. (%) (1-shot)
ImageNet	46.6±4.3
UCDS [2]	31.1±3.5
CDS	**66.8**±2.1

Table H: Comparison with [2] when finetuned with source 1-shot label per class in the Office D→A setting using CDAN. [2] hurts the performances of the ImageNet pre-trained weights.

Adapt	Pre-train	CUB: Target Acc (%) on 1-shot/3-shots		
		Real→Painting	Painting→Real	AVG
SO	ImageNet	5.8±0.7 / 18.5±0.6	4.4±0.3 / 11.8±0.4	5.1 / 15.0
	CDS	**29.7**±1.6 / **43.9**±0.7	**11.8**±0.4 / **22.9**±0.7	**20.8 / 33.4**
DANN	ImageNet	8.4±0.4 / 22.4±1.0	3.8±0.2 / 12.8±1.0	6.1 / 17.6
	CDS	**28.2**±1.5 / **44.3**±0.1	**12.2**±0.2 / **25.0**±1.4	**20.2 / 34.6**
CDAN	ImageNet	7.4±1.1 / 22.1±0.5	5.7±0.6 / 14.9±0.5	6.5 / 18.5
	CDS	**31.8**±1.4 / **47.2**±0.7	**14.7**±0.2 / **29.8**±1.9	**23.2 / 38.5**
MME	ImageNet	12.5±0.5 / 45.9±0.8	11.6±1.0 / 37.9±1.1	12.0 / 41.9
	CDS	**35.1**±0.9 / **50.3**±0.6	**22.3**±1.0 / **44.5**±1.7	**28.7 / 47.4**

Table I: Detailed results of Table 6 on CUB. Target accuracy (%) on few-shot domain adaptation with source 1-shot and 3-shots labels per class. The second column (Pre-train) refers to pre-training methods used in these experiments.

Adapt	Pre-train	Office-Home: Target Acc. (%)													
		Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	AVG Acc.	AVG Std.
(a) 1-shot															
SO	ImageNet	12.9	18.6	22.9	8.5	10.1	11.7	18.7	17.4	30.1	27.1	17.1	29.5	18.7	1.3
	CDS	**23.3**	**34.3**	**40.7**	**24.9**	**26.2**	**30.1**	**36.7**	**28.5**	**47.2**	**40.8**	**29.3**	**44.1**	**33.8**	1.4
DANN	ImageNet	12.1	19.2	21.6	10.4	12.0	12.7	21.8	17.0	32.6	27.1	20.2	34.2	20.1	1.6
	CDS	**26.2**	**38.4**	**43.9**	**25.4**	**27.5**	**31.0**	**36.5**	**29.8**	**47.9**	**39.6**	**32.9**	**46.9**	**35.5**	1.9
CDAN	ImageNet	12.8	20.5	23.3	10.3	11.7	13.1	20.0	17.4	31.5	27.7	17.7	29.7	19.6	1.3
	CDS	**26.3**	**36.2**	**42.7**	**26.2**	**26.6**	**32.2**	**37.2**	**29.9**	**46.7**	**39.5**	**32.3**	**44.8**	**35.0**	1.2
MME	ImageNet	22.5	26.1	29.0	15.3	16.5	15.2	32.9	30.9	43.3	39.6	33.5	41.5	28.8	1.6
	CDS	**28.1**	**32.0**	**39.2**	**24.1**	**28.0**	**33.2**	**38.7**	**31.6**	**51.4**	**42.7**	**38.5**	**47.7**	**36.3**	2.1
SRDC	ImageNet	16.0	30.3	35.3	16.0	19.1	20.6	30.1	23.3	42.1	37.0	26.3	41.9	28.6	1.5
	CDS	**26.6**	**41.4**	**47.5**	**36.0**	**34.5**	**37.9**	**42.7**	**33.3**	**55.1**	**47.5**	**38.7**	**54.5**	**41.3**	1.7
(b) 3-shots															
SO	ImageNet	23.4	36.5	45.7	15.5	21.9	22.2	31.6	23.3	49.8	40.8	27.5	48.4	32.4	1.5
	CDS	**30.3**	**46.7**	**55.4**	**35.0**	**41.9**	**43.6**	**46.0**	**35.8**	**62.7**	**53.9**	**37.9**	**59.2**	**45.7**	1.3
DANN	ImageNet	25.3	39.4	46.3	17.8	24.7	25.6	33.1	25.1	51.2	40.5	29.5	52.4	34.2	1.2
	CDS	**35.4**	**51.3**	**58.3**	**37.0**	**41.8**	**46.3**	**48.4**	**38.4**	**64.4**	**54.7**	**43.9**	**63.4**	**48.6**	1.2
CDAN	ImageNet	24.2	38.5	48.0	18.7	26.5	28.7	34.6	27.4	55.5	44.6	29.7	45.3	35.0	1.8
	CDS	**37.1**	**50.1**	**59.4**	**42.1**	**46.5**	**50.1**	**51.1**	**41.7**	**66.2**	**58.2**	**45.8**	**64.9**	**51.1**	1.6
MME	ImageNet	40.5	47.9	53.9	39.1	42.4	44.3	50.5	44.8	64.5	59.6	49.7	66.7	50.3	5.1
	CDS	**44.3**	**53.6**	**61.0**	**47.6**	**50.7**	**55.5**	**55.1**	**46.0**	**67.7**	**61.7**	**51.6**	**67.9**	**55.2**	5.4
SRDC	ImageNet	39.5	48.7	53.2	40.0	42.5	45.8	51.9	40.0	66.9	60.6	42.8	64.5	48.9	1.6
	CDS	**36.9**	**55.4**	**62.8**	**50.2**	**55.3**	**56.4**	**56.0**	**42.0**	**71.2**	**64.6**	**50.6**	**69.3**	**55.9**	1.6

Table J: Detailed results of Table 6 on Office-Home. Target accuracy (%) on few-shot domain adaptation with source 1-shot and 3-shots labels per class on Office-Home.

Adapt	Pre-train	Office: Target Acc. (%) on 1-shot / 3-shots							
		A→D	A→W	D→ A	D→W	W→A	W→D	AVG Acc.	AVG Std.
SO	ImageNet	29.7 / 46.3	32.7 / 49.7	39.4 / 54.8	51.2 / 84.0	33.4 / 51.5	37.4 / 85.4	37.3 / 61.9	3.6 / 2.3
	CDS	**48.3 / 65.9**	**49.2 / 65.5**	**61.4 / 64.4**	**77.5 / 90.4**	**57.4 / 64.4**	**71.5 / 93.0**	**60.9 / 73.9**	1.7 / 2.2
DANN	ImageNet	37.6 / 53.7	35.6 / 53.5	46.8 / 56.1	72.0 / 86.6	43.5 / 55.0	67.1 / 87.2	50.4 / 65.3	3.2 / 2.1
	CDS	**50.6 / 65.4**	**53.4 / 67.1**	**62.9 / 67.1**	**78.0 / 90.2**	**60.1 / 66.8**	**73.8 / 91.0**	**63.1 / 74.6**	2.7 / 2.5
CDAN	ImageNet	36.0 / 58.3	36.8 / 65.7	46.6 / 67.3	60.4 / 91.5	40.6 / 67.7	55.4 / 93.6	46.0 / 74.0	4.2 / 3.8
	CDS	**52.4 / 72.4**	**55.0 / 74.2**	**66.8 / 73.5**	**79.2 / 92.5**	**62.5 / 67.8**	**75.2 / 94.8**	**65.2 / 79.2**	2.1 / 2.8
SRDC	ImageNet	47.2 / 64.1	48.7 / 69.2	62.3 / 70.5	80.4 / 92.5	53.9 / 67.6	70.8 / 90.2	60.5 / 75.7	3.0 / 2.4
	CDS	**59.1 / 74.3**	**57.8 / 72.8**	**68.8/ 73.1**	**85.7 / 92.6**	**63.9 / 72.2**	**79.7 / 93.6**	**69.2 / 79.8**	3.9 / 2.4
MME	ImageNet	43.2 / 63.7	47.6 / 68.0	58.6 / 67.5	78.7 / 91.9	53.9 / 64.5	72.6 / 92.0	59.1 / 74.6	3.3 / 2.6
	CDS	**51.3 / 74.9**	**55.8 / 73.4**	**65.0/ 70.2**	**86.0 / 93.1**	**60.6 / 69.7**	**75.8/ 95.5**	**65.8 / 79.5**	3.1 / 1.9

Table K: Detailed results of Table 6 on Office. Target accuracy (%) on few-shot domain adaptation with source 1-shot and 3-shots labels per class. CDS improves accuracy of diverse DA methods in all settings compared to ImageNet weights.