This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Link-based Contrastive Learning for One-Shot Unsupervised Domain Adaptation

Yue Zhang Mingyue Bin YuyangZhang Zhongyuan Wang Zhen Han Chao Liang*
School of Computer Science, Wuhan University
National Engineering Research Center for Multimedia Software(NERCMS)
Hubei Key Laboratory of Multimedia and Network Communication Engineering, Wuhan
{moozy924, bin_mingyue, 2021302131204, hanzhen, cliang}@whu.edu.cn, wzy_hope@163.com

Abstract

Unsupervised domain adaptation (UDA) transfers knowledge from a labeled source domain to an unlabeled target domain via distribution alignment. However, in real-world scenarios like public safety or access control, obtaining sufficient source data is challenging, limiting existing UDA methods. This paper investigates a realistic but rarely studied problem called one-shot unsupervised domain adaptation (OSUDA), where only one source example per category is available. OSUDA faces dual challenges in feature learning and domain alignment due to the extreme source data scarcity. To address these, we propose link-based contrastive learning (LCL), a simple yet effective approach for OSUDA. LCL leverages in-domain links to learn discriminative features from abundant unlabeled target data and cross-domain links to achieve precise domain alignment with only one source sample per category. Extensive experiments on four domain adaptation benchmarks (VisDA-2017, Office-31, Office-Home, and DomainNet) demonstrate LCL's effectiveness under the OSUDA setting. Additionally, we construct a real-world OSUDA surveillance face recognition dataset, where LCL consistently improves recognition performance across various face recognition methods.

1. Introduction

Deep learning has achieved remarkable performance in the past decade. But due to the *domain shift* problem [11], deep neural networks trained on specific datasets often fail to generalize to new domains. Unsupervised domain adaptation (UDA) methods [17, 18, 30, 47, 55] address this issue by transferring knowledge from a labeled source domain to an unlabeled target domain. Conventional UDA methods assume abundant labeled source and unlabeled target data, focusing on learning discriminative knowledge from the source and aligning data distributions across domains.

Figure 1. A real-world cross-domain surveillance video face recognition task (eyes are obscured for the sake of privacy protection.). The source domain contains only one clear registration portrait for each identity, and the target domain contains abundant blurry and angled faces captured by surveillance cameras.

In real-world scenarios, however, abundant source data can be hard to obtain, leading to new research directions. Few-shot UDA (FUDA) methods [22, 52] address cases with very few labeled source samples but abundant unlabeled source data for self-supervised learning and distribution estimation. Source-free DA (SFDA) methods [21, 28, 51] handle scenarios where source data are unavailable during adaptation, using a model pretrained on fully-supervised source data to approximate discriminative knowledge and source domain distributions.

This paper studies a realistic but more challenging setting called one-shot UDA (OSUDA), where *the source domain is composed of only one example from each category* and *the target domain contains abundant unlabeled samples*.[1] Take a real-world surveillance video face recognition task as an example: as shown in Figure 1, each person has one clear frontal face image (source domain) and many blurry angled faces from high-mounted cameras (target do-

*Chao Liang is the corresponding author.

[1] In the current OSUDA problem, we confine our study to the simplest close-set setting [4], in which the source and target domain share the same set of category labels.

main). To protect privacy, obtaining extra high-quality portraits for FUDA or a pretrained model for SFDA is impractical. Therefore, OSUDA faces tougher dual challenges: (1) the scarcity of source samples hinders discriminative feature learning, and (2) the quantity gap between source and target data complicates domain alignment. These challenges will be further discussed in the Related Work.

To this end, we propose a link-based contrastive learning (LCL) method for OSUDA. To address the first challenge, LCL performs contrastive learning based on in-domain links showing whether two samples are from the same cluster. By directly aggregating samples in the same cluster with *must-links* and dispersing samples in different clusters with *cannot-links*, LCL is able to learn discriminative features with abundant unlabeled target samples. And to address the second challenge, LCL uses cross-domain links for contrastive learning, pulling bidirectionally matched clusters with *must-links* and pushing apart mismatched ones with *cannot-links*, enabling knowledge transfer by aligning semantically consistent clusters across domains.

To summarize, our contributions are three-fold:

- We investigate a challenging yet insufficiently studied UDA setting called OSUDA, where *the source domain is composed of only one example from each category* and *the target domain contains massive unlabeled samples*.
- We propose LCL to address the OSUDA problem. LCL generates in-domain links for discriminative feature learning and constructs cross-domain links to achieve domain alignment.
- We validate the LCL method on 4 public domain adaptation benchmarks and a self-built real-world OSUDA surveillance video face recognition task.

2. Related Work

Unsupervised domain adaptation. UDA aims to transfer knowledge from a fully-labeled source domain to an unlabeled target domain. Most UDA methods focus on aligning data distribution in the feature space. A straightforward way is to minimize statistical discrepancy between the source and the target domain [20, 40, 46], where Maximum Mean Discrepancy (MMD) [14] is one of the most commonly used metrics. With the development of Generative Adversarial Networks (GAN) [13], some methods [10, 35, 38, 54] propose to train a feature encoder and a domain discriminator in an adversarial manner. [37] and [3] further remove the discriminator, playing the min-max game between feature encoder and category classifier. However, UDA methods rely heavily on abundant labeled source data, making them incompatible with OSUDA's scarcity of labeled data. *e.g.*, as illustrated in Figure 2, the classification accuracy of SDAT [35], one of the SOTA UDA methods, significantly drops from over 75% to 39.36% as the source data volume decreases from 100% to 1-shot per category.

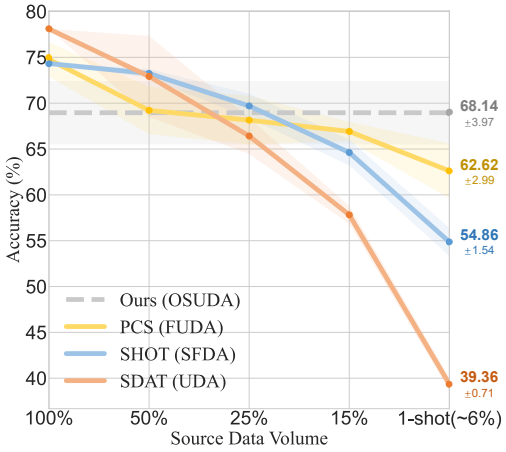


Figure 2. The dynamic performance of existing methods with decreasing source data volume on the D→A task in Office-31 dataset. The solid line and shading represent the mean accuracy and standard deviation of 5 repeated experiments, respectively.

Few-shot unsupervised domain adaptation. FUDA is relevant research topic that studies the case where source data are abundant but with limited labels. In this scenario, existing works focus on learning discriminative features in a self-supervised manner. CDS [22] performs instance-level contrastive learning to learn discriminative features and minimizes the matching entropy to align confidently matched sample pairs across domains. PCS [52] replaces the instance-level learning in CDS with cluster-level learning to make the feature more semantic-aware. UCDIR [19] adopts cluster-level contrastive learning to learn semantic-aware features but proposes a novel distance-to-distance loss to minimize the domain discrepancy. The success of contrastive learning in these FUDA methods provides a promising solution for the OSUDA problem to learn domain-invariant discriminative features without any external supervision. But the scarcity of source data in OSUDA still hinders FUDA methods from mining latent knowledge from the source domain. We plot the dynamic performance of PCS [52] with decreasing size of unlabeled source data in Figure 2, where the number of labeled source data remains 1-shot per category. And the result shows that the classification accuracy drops by over 10% as the size of unlabeled data decreases from 100% to 0.

Source-free domain adaptation. SFDA is another relevant research topic. It investigates the situation where one can only access abundant unlabeled target data and a model pretrained by fully-labeled source data during the adaptation process. SHOT [27] generates pseudo-labels for target samples with the fixed source classifier to train the model. Since then, a series of methods [28, 48–50] have focused on mining the neighbor relationship between target samples to improve the reliability of pseudo-labeling. Other ap-

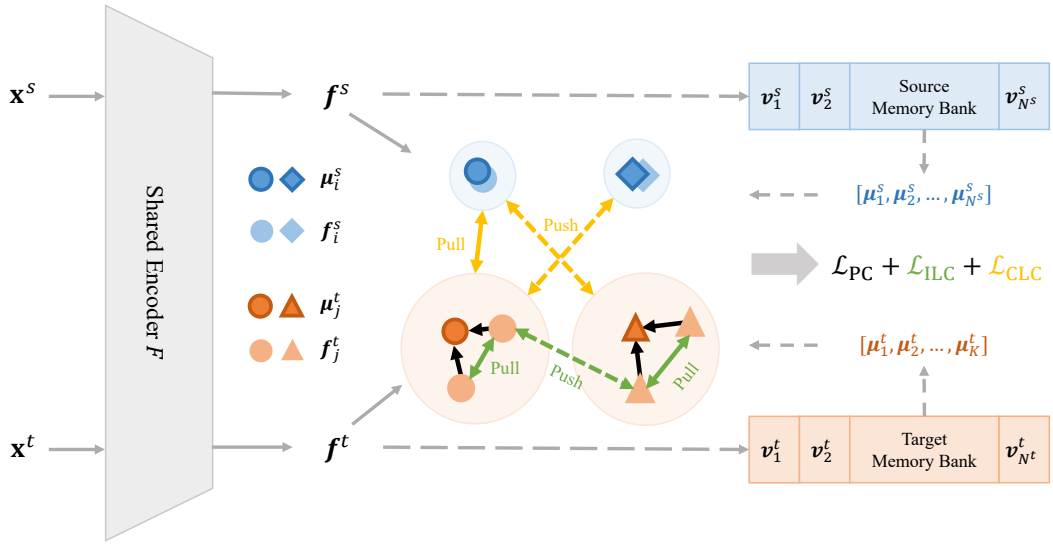


Figure 3. The overview of the proposed LCL, where $\mathbf{x}$, $\boldsymbol{f}$, and $\boldsymbol{\mu}$ represent input images, normalized features, and prototypes, respectively. $\mathcal{L}_{\text{PC}}$ encourages samples to approach their prototypes (black solid lines) to learn discriminative features. $\mathcal{L}_{\text{ILC}}$ reinforces the learning process by pulling together samples in the same cluster (green solid lines) and pushing apart those in different clusters (green dashed line). To align semantics across domains, $\mathcal{L}_{\text{CLC}}$ pulls together bidirectionally matched clusters (yellow solid line) and pushes apart mismatched clusters (yellow dashed lines).

proaches [9, 26, 45] aim to apply techniques from the UDA method, including adversarial learning, minimizing statistical metrics across domains, *etc.*, to the SFDA problem. In general, SFDA methods depend on well-pretrained source models to provide source knowledge, thus they tend to be powerless in the OSUDA setting where source model cannot be fully pretrained. We test the performance of SHOT [27], one of the SOTA SFDA methods, with different source models pretrained by decreasing source data volume. And we can observe from Figure 2 that its performance is almost the same as that of SDAT when the model is fully pretrained in the source domain, but considerably drops as the source data volume involved in the pretraining stage decreases.

Summary. In all, OSUDA is a more challenging task compared with existing settings. (1) Compared with UDA, OSUDA lacks abundant labeled source data to learn discriminative features and to accurately estimate source distributions for domain alignment. (2) Compared with FUDA, OSUDA has no access to abundant unlabeled source data to facilitate the adaptation process with extra latent knowledge. (3) Compared with SFDA, OSUDA cannot access the well-pretrained source model to obtain source knowledge. These issues hinder the direct application of existing methods in the OSUDA setting, thus motivate us to explore a tailored solution for the new problem.[2]

[2]It is worth noting that the name "OSUDA" has also appeared in [12, 31, 44], where the target domain contains only one sample per class, unlike the source domain studied in this paper. This difference poses unique challenges for knowledge transfer and domain adaptation.

3. Method

3.1. Preliminaries

In the OSUDA setting, the source domain $\mathcal{D}_s = \{(\mathbf{x}_i^s, y_i^s)\}_{i=1}^{N^s}$ is composed of only one labeled sample from each category, *i.e.*, $y_i \neq y_j$ if $i \neq j$. And the target domain $\mathcal{D}_t = \{\mathbf{x}_i^t\}_{i=1}^{N^t}$ consists of abundant unlabeled samples sharing the same label set as their source counterparts. Since one can directly categorize a target sample by calculating the similarity between its feature and all source features in the OSUDA setting, we aim to train a domain-invariant discriminative feature encoder F on both $\mathcal{D}_s$ and $\mathcal{D}_t$, and evaluate the classification accuracy on $\mathcal{D}_t$. The overview of the proposed method is illustrated in Figure 3.

3.2. In-domain Feature Learning

In order to learn semantic-aware discriminative features, we follow PCS [52] to perform prototypical contrastive learning within the target domain.[3] Specifically, we maintain a memory bank, $\mathbf{V}^t = [\boldsymbol{v}_1^t, \cdots, \boldsymbol{v}_{N^t}^t]$, to store the gradient-detached target features, where $\boldsymbol{v}_i^t$ is initialized with $\boldsymbol{f}_i^t$ and updated with a momentum $\lambda \in [0, 1]$ after each epoch by $\boldsymbol{v}_i^t \leftarrow \lambda \boldsymbol{v}_i^t + (1 - \lambda)\boldsymbol{f}_i^t$. At the beginning of each epoch, K-means clustering is performed on $\mathbf{V}^t$ to get K clusters $\mathcal{C}^t = \{C_1^t, \cdots, C_K^t\}$ with normalized semantic prototypes, *i.e.*, cluster centroids, $\{\boldsymbol{\mu}_i^t\}_{i=1}^K$. In order to keep the logical

[3]Since our OSUDA setting provides abundant data only in the target domain, the learning process is restricted to the target domain.

symmetry across domains, we maintain a similar memory bank $\mathbf{V}^s$ in the source domain. And since there is only one sample for each category, we directly set $C_i^s = \{\mathbf{x}_i^s\}$ and $\boldsymbol{\mu}_i^s = \boldsymbol{v}_i^s$, where $i = 1, \cdots, N^s$. The objective for prototypical contrastive learning is given by

$$\mathcal{L}_{\text{PC}} = -\frac{1}{N^t}\sum_{i=1}^{N^t}\sum_{j=1}^{K} p_{ij}\log\frac{\exp(\boldsymbol{f}_i^t\cdot\boldsymbol{\mu}_j^t)}{\sum_{k=1}^{K}\exp(\boldsymbol{f}_i^t\cdot\boldsymbol{\mu}_k^t)} \tag{1}$$

where $p_{ij} = 1$ if $\mathbf{x}_i^t \in C_j^t$, and $p_{ij} = 0$ otherwise.

However, PCS is designed for FUDA problems where discriminative features can be learned from both domains, yet it's impractical to mine discriminative information with only one sample per category from the source domain in the OSUDA setting. Therefore, we attempt to excavate more discriminative information from the target domain. In addition to $\mathcal{L}_{\text{PC}}$, which use sample-prototype relationship, we construct pairwise sample links and use contrastive learning to attract samples within the same cluster and disperse those in different clusters. Such an attracting-and-dispersing mechanism may seem similar to that in AaD [50]. But unlike AaD that constrains label predictions with feature similarities, in-domain links constrain feature similarities using clustering results, while cross-domain links, introduced in the next section, use bidirectional matching results to constrain feature similarities.Specifically, the in-domain link q_{ij} between a pair of target samples $(\boldsymbol{f}_i^t, \boldsymbol{f}_j^t)$ is denoted by

$$q_{ij} = \begin{cases} +1, \text{ if } P^t(i) = P^t(j), \\ -1, \text{ otherwise.} \end{cases} \tag{2}$$

where $P^t(i)$ returns the index of the cluster where $\mathbf{x}_i^t$ belongs, $q_{ij} = +1$ and $q_{ij} = -1$ represent the *must-link* and *cannot-link* constraints between a pair of target samples, respectively.

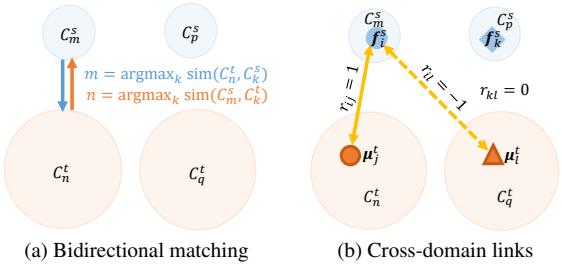
Thereafter, the in-domain link-based contrastive loss is

$$\begin{aligned}\mathcal{L}_{\text{ILC}} = &-\frac{1}{N^t}\sum_{i=1}^{N^t}\sum_{j=1}^{N^t}\frac{q_{ij}+1}{2}\log\frac{\exp(\boldsymbol{f}_i^t\cdot\boldsymbol{f}_j^t)}{\sum_{k=1}^{N^t}\exp(\boldsymbol{f}_i^t\cdot\boldsymbol{f}_k^t)} \\ &+\frac{1}{N^t}\sum_{i=1}^{N^t}\sum_{j=1}^{N^t}\frac{q_{ij}-1}{2}\log\frac{\exp(1-\boldsymbol{f}_i^t\cdot\boldsymbol{f}_j^t)}{\sum_{k=1}^{N^t}\exp(1-\boldsymbol{f}_i^t\cdot\boldsymbol{f}_k^t)}\end{aligned} \tag{3}$$

where the first term pulls close samples in the same cluster via the *must-links*, and the second term pushes away samples in different clusters through the *cannot-links*.

3.3. Cross-domain Semantic Alignment

Due to the sparse source data, estimating the source distribution or to constructing a robust domain classifier in OSUDA is impractical. Consequently, methods minimizing statistical discrepancy may misalign semantic structures and adversarial methods often fail to produce domain-invariant features. Recent FUDA methods [22, 52] use entropy minimization to match each sample to its nearest counterpart or cluster in the other domain. But due to the domain shift, cross-domain similarity measurement is often unreliable, leading to suboptimal matching results (see Tables 1-4).



(a) Bidirectional matching (b) Cross-domain links

Figure 4. Cross-domain semantic alignment in LCL.

To address this issue, we propose bidirectional matching between source and target clusters, then construct links between each sample and prototypes of the other domain based on the bidirectional matching results of the clusters they belong. The bidirectional matching mechanism enforces stricter conditions, and the cluster-level matching aggregates results across semantically similar samples, leading to more reliable matching in LCL.

Specifically, as illustrated in Figure 4(a), a cluster pair (C_m^s, C_n^t) is considered to be bidirectionally matched *iff* it is simultaneously satisfied that $n = \text{argmax}_k\text{sim}(C_m^s, C_k^t)$ and $m = \text{argmax}_k\text{sim}(C_n^t, C_k^s)$, where $\text{sim}(C_m^s, C_n^t) = \boldsymbol{\mu}_m^s \cdot \boldsymbol{\mu}_n^t$.

Next, we set links between samples and prototypes of the other domain based on the above bidirectional matching results. Without loss of generality, we first discuss the matching process from the source domain to the target domain. Specifically, as illustrated in Figure 4(b), a *must-link* (denoted by the yellow solid line) is generated between $(\boldsymbol{f}_i^s, \boldsymbol{\mu}_j^t)$ as their corresponding clusters C_m and C_n are bidirectionally matched. On this basis, we further assume that it is illegal to match an already bidirectionally matched cluster with any other clusters, hence we treat (C_m^s, C_q^t) as an illegally matched cluster pair and establish a *cannot-link* (denoted by the yellow dashed line) between $(\boldsymbol{f}_i^s, \boldsymbol{\mu}_l^t)$. As for C_p^s, since it is not bidirectionally matched with any clusters in the target domain, we consider it as a semantically uncertain cluster and won't set any links for its sample $\boldsymbol{f}_k^s$.

Table 1. Acc. (%) on VisDA-2017.

Category	Method	Venue	Acc.
w/o. DA	Direct	Baseline	40.20
UDA/FUDA	PCS	CVPR 2021	56.22
	UCDIR	ECCV 2022	45.60
	SDAT	ICML 2022	49.67
	CAF-A	TKDE 2023	34.55
	CAF-D	TKDE 2023	50.28
	ICON	NeurIPS 2023	55.46
	ProtoOT	AAAI 2024	34.38
	MCUDA	TOMM 2024	49.77
SFDA	SHOT	ICML 2020	59.67
	NRC	NeurIPS 2021	38.94
	AaD	NeurIPS 2022	39.81
	SFDA-DE	CVPR 2022	59.23
	BMD	ECCV 2022	61.11
	GPUE	CVPR 2023	55.92
	SFADA	PR 2023	49.24
	TPDS	IJCV 2024	61.31
OSUDA	LCL	Ours	**63.09**

Table 2. Adaptation accuracy (%) on Office-31.

Method	A→D	A→W	D→A	D→W	W→A	W→D	Avg.
Direct	55.26	54.42	53.46	77.41	50.49	75.42	61.08
PCS	65.98	63.34	62.62	**86.44**	61.72	84.74	70.81
UCDIR	34.26	35.95	38.48	54.49	34.15	45.70	40.50
SDAT	39.60	32.88	39.36	63.60	32.72	56.54	44.12
CAF-A	57.47	58.31	54.30	83.22	50.37	79.52	63.87
CAF-D	49.72	53.13	52.47	79.22	47.77	77.75	59.98
ICON	63.67	68.67	66.26	85.96	**67.68**	78.58	71.80
ProtoOT	38.96	39.12	39.63	62.87	33.28	57.79	45.28
MCUDA	40.22	33.42	39.36	63.60	32.72	56.54	44.31
SHOT	54.54	51.09	54.86	73.33	51.35	70.72	59.32
NRC	51.73	47.52	46.86	72.05	43.29	73.25	55.78
AaD	51.73	47.52	47.18	72.05	43.59	73.25	55.89
SFDA-DE	60.84	61.54	57.01	78.39	52.86	75.82	64.41
BMD	51.57	48.38	53.49	72.43	52.69	74.05	58.77
GPUE	61.29	58.59	54.53	78.93	50.54	75.50	63.23
SFADA	54.86	55.25	56.35	75.02	51.55	75.46	61.42
TPDS	57.47	57.59	52.33	71.13	51.64	75.34	60.92
LCL	**69.28**	**68.96**	**68.14**	84.10	65.46	**90.12**	**74.34**

The cross-domain link r_{ij} between $(\boldsymbol{f}_i^s, \boldsymbol{\mu}_j^t)$ is denoted by

$$r_{ij} = \begin{cases} +1, \text{ if } R^{s\to t}(P^s(i)) = j, \\ -1, \text{ if } R^{s\to t}(P^s(i)) \neq -1 \text{ and } R^{s\to t}(P^s(i)) \neq j, \\ 0, \text{ otherwise.} \end{cases} \tag{4}$$

where $R^{s\to t}(\cdot)$ returns the index of the bidirectionally matched cluster in the target domain given a cluster index in the source domain, and returns -1 if the given source cluster is semantically uncertain. $r_{ij} = 0$ means no links for the given pair, $r_{ij} = +1$ and $r_{ij} = -1$ denote the *must-link* and *cannot-link* constraints for the bidirectionally matched and illegally matched pairs, respectively.

On this basis, we perform cross-domain link-based contrastive loss by

$$\begin{aligned} \mathcal{L}_{\text{CLC}}^{s\to t} = &- \frac{1}{N^s} \sum_{i=1}^{N^s} \sum_{j=1}^{K} \left[\frac{r_{ij}+1}{2} \cdot r_{ij} \right] \log \frac{\exp(\boldsymbol{f}_i^s \cdot \boldsymbol{\mu}_j^t)}{\sum_{k=1}^{K} \exp(\boldsymbol{f}_i^s \cdot \boldsymbol{\mu}_k^t)} \\ &- \frac{1}{N^s} \sum_{i=1}^{N^s} \sum_{j=1}^{K} \left[\frac{r_{ij}-1}{2} \cdot r_{ij} \right] \log \frac{\exp(1 - \boldsymbol{f}_i^s \cdot \boldsymbol{\mu}_j^t)}{\sum_{k=1}^{K} \exp(1 - \boldsymbol{f}_i^s \cdot \boldsymbol{\mu}_k^t)} \end{aligned} \tag{5}$$

where the first term pulls close samples and prototypes whose corresponding clusters are bidirectionally matched, and the second term pushes away samples and prototypes whose corresponding clusters are illegally matched.

Similarly, we calculate $\mathcal{L}_{\text{CLC}}^{t\to s}$ to conduct cross-domain link-based contrastive learning between target samples and source prototypes, and the overall loss function is

$$\mathcal{L}_{\text{CLC}} = \mathcal{L}_{\text{CLC}}^{s\to t} + \mathcal{L}_{\text{CLC}}^{t\to s} \tag{6}$$

3.4. LCL for OSUDA

The LCL optimizes $\mathcal{L}_{\text{PC}}$ and $\mathcal{L}_{\text{ILC}}$ to learn discriminative features, and minimizes $\mathcal{L}_{\text{CLC}}$ to align semantically consistent clusters across domains. The overall loss function is

$$\mathcal{L}_{\text{LCL}} = \mathcal{L}_{\text{PC}} + \alpha \mathcal{L}_{\text{ILC}} + \beta \mathcal{L}_{\text{CLC}} \tag{7}$$

where α and β are hyperparameters to balance link-based contrastive learning loss terms. In the evaluation process, the prediction of $\mathbf{x}_i^t$ is obtained by $\hat{y}_i^t = \text{argmax}_j(\boldsymbol{f}_i^t \cdot \boldsymbol{v}_j^s)$.

4. Experiments

4.1. Settings

Datasets. We evaluate our method on 4 public datasets. **VisDA-2017** [32] is a challenging simulation-to-real dataset that contains over 280,000 images of 12 categories. **Office-31** [36] is a real-world domain adaptation dataset with 3 domains (Amazon, DSLR, Webcam) in 31 classes. **Office-Home** [42] is a more difficult dataset that contains 4 domains with 65 classes. **DomainNet** [33] is a large-scale domain adaptation benchmark. Following [52], we adopt a subset that contains 4 domains (Clipart, Real, Painting, Sketch) with 126 classes. In particular, we construct a real-world one-shot cross-domain facial recognition dataset called **VSA**, with 46 high-quality portraits in the source domain and 3,803 low-quality faces from 6 surveillance cameras in the target domain. The examples of the VSA dataset are shown in Figure 1. More details and experimental results about the VSA dataset are presented in Appendices A.

Implementation Details. We perform our experiments on a server with a 16 GB NVIDIA V100 GPU and a 128 GB RAM. For 4 public datasets, we adopt ResNet-50 [16] pretrained on ImageNet [5] as our backbone. And

Table 3. Adaptation accuracy (%) comparison on Office-Home.

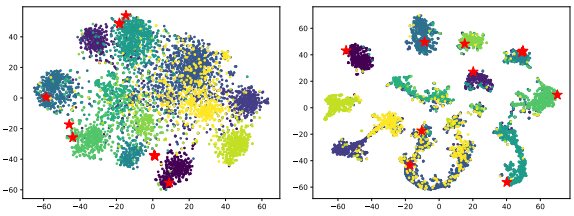
Method	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg.
Direct	18.25	30.42	36.65	24.72	27.02	29.11	31.13	22.81	44.76	35.58	23.00	41.90	30.45
PCS	26.62	42.62	49.36	33.68	40.68	41.84	38.47	30.09	56.97	44.25	29.86	54.47	40.74
UCDIR	15.64	24.55	30.20	17.12	22.68	22.63	24.94	18.14	33.10	28.68	18.86	37.03	24.46
SDAT	13.40	19.92	25.62	15.13	15.89	16.89	22.17	15.46	31.62	29.55	16.49	31.23	21.11
CAF-A	15.02	32.12	40.13	17.87	22.18	24.46	26.44	17.32	42.47	34.12	20.43	47.00	28.30
CAF-D	15.85	26.48	33.92	18.67	20.26	22.31	27.22	19.39	40.69	35.02	20.75	42.04	26.88
ICON	23.64	33.30	45.82	30.82	28.10	31.68	38.70	28.46	53.58	44.85	28.98	49.58	36.44
ProtoOT	14.66	22.52	27.52	17.42	21.51	22.63	24.05	16.55	32.24	29.17	17.25	34.35	23.32
MCUDA	16.20	20.92	27.22	15.53	18.89	19.09	21.21	17.46	30.94	31.35	16.61	31.95	22.28
SHOT	21.53	31.28	37.21	24.23	26.91	28.73	32.44	24.64	45.12	37.42	26.46	45.85	31.82
NRC	19.33	29.02	25.56	22.37	24.56	26.77	30.71	23.36	42.36	35.12	24.32	40.64	29.51
AaD	19.33	29.02	35.36	22.37	24.56	26.77	30.71	23.36	42.37	35.12	24.32	40.64	29.51
SFDA-DE	24.19	34.07	40.33	26.25	29.11	30.96	34.07	29.08	47.72	38.48	27.99	46.57	34.07
BMD	21.91	34.78	41.12	25.92	30.52	32.07	34.72	26.65	47.97	37.67	26.79	46.47	33.88
GPUE	19.99	33.03	40.84	24.11	27.06	28.07	35.48	23.56	47.16	41.14	25.55	47.90	32.82
SFADA	19.26	28.67	35.71	24.81	23.27	27.65	34.53	25.45	45.76	37.55	25.52	44.47	31.05
TPDS	25.13	33.71	40.99	26.66	28.45	31.64	34.91	29.90	43.14	38.70	26.85	45.17	33.77
LCL	**32.55**	**49.57**	**55.91**	**38.88**	**46.18**	**46.95**	**48.17**	**36.43**	**61.03**	**51.77**	**37.58**	**61.99**	**47.25**

Table 4. Adaptation accuracy (%) comparison on DomainNet.

Method	Cl→Pa	Cl→Re	Cl→Sk	Pa→Cl	Pa→Re	Pa→Sk	Re→Cl	Re→Pa	Re→Sk	Sk→Cl	Sk→Pa	Sk→Re	Avg.
Direct	11.24	17.45	9.11	10.26	28.88	10.24	14.27	26.79	14.18	9.97	13.45	18.22	15.34
PCS	15.67	25.78	10.41	10.37	37.87	10.76	17.42	30.62	14.18	14.16	19.88	**28.78**	19.99
UCDIR	12.02	15.39	13.95	15.22	23.97	20.59	18.78	33.85	28.19	15.57	20.64	25.07	21.10
SDAT	8.53	14.36	7.36	10.28	23.51	8.30	14.54	22.72	11.53	9.44	9.40	13.95	12.83
CAF-A	8.96	18.36	6.78	12.35	29.16	8.17	18.08	23.77	12.67	8.63	8.93	16.97	14.40
CAF-D	10.61	16.18	6.37	11.24	25.30	7.92	15.58	25.69	11.55	9.30	10.64	15.96	13.86
ICON	8.78	13.98	6.24	9.34	25.62	6.80	15.27	21.36	9.66	7.70	6.51	11.18	11.87
ProtoOT	8.26	13.57	9.10	8.19	25.84	12.12	11.59	26.25	14.87	9.84	14.24	18.57	14.37
MCUDA	11.75	15.55	8.89	11.33	27.65	9.98	14.80	25.10	15.38	10.71	11.99	16.64	14.98
SHOT	10.17	16.49	7.74	11.44	29.66	9.29	16.60	27.62	14.59	8.53	12.35	18.60	15.26
NRC	10.32	16.59	10.16	14.41	25.42	12.95	20.67	26.80	19.40	11.36	13.07	17.68	16.54
AaD	10.31	16.55	10.16	14.41	27.36	12.93	20.69	26.74	19.00	11.40	13.03	17.58	16.68
SFDA-DE	10.42	17.81	11.27	16.26	30.72	14.88	23.11	30.67	22.36	12.02	12.84	17.91	18.35
BMD	11.29	21.31	10.63	16.39	34.91	14.05	22.77	29.52	20.35	13.24	14.93	21.83	19.27
GPUE	13.28	18.24	13.43	18.05	31.38	19.52	26.16	35.10	27.50	13.95	17.39	21.13	21.62
SFADA	5.76	11.34	5.10	8.46	18.37	6.42	13.07	18.77	8.41	7.19	8.67	12.90	10.37
TPDS	10.65	16.39	8.02	11.96	31.36	9.21	16.14	26.84	15.05	9.53	11.77	19.72	15.55
LCL	**20.44**	**28.42**	**15.78**	**23.38**	**44.21**	**21.77**	**31.29**	**44.24**	**28.19**	**20.12**	**25.84**	28.16	**27.65**

for VSA, we adopt six popular face recognition backbones based on the ResNet-50 architecture, including CosFace [43] pretrained on Glint30K [2], PartialFC [2] pretrianed on Glint360k [1], AdaFace [23] pretrained on MS1MV2 [6], ArcFace [6], SphereFace [29] and VPL [8] pretrained on MS1MV3 [7], and compare recognition performances using features learned by various domain adaptation methods. L2-normalization is performed on the output features. And we use an SGD optimizer with a momentum of 0.9, a learning rate of 0.01, and a batch size of 128. We choose $\lambda = 0.5$ as our default hyperparameter. Following [52], we perform K-means clustering multiple times to mitigate random effects, and set $K = 2N^s$ to generate more purified clustering results. On the four public datasets, we empirically set $\alpha = \beta = 1$. And on the VSA dataset, we set $\alpha = 0.5$, $\beta = 5$. The detailed analysis of λ and K, α and β are presented in Appendix B.1.

Comparative Baselines. Since no method is specifically designed for OSUDA, we compare LCL with state-of-the-art (SOTA) UDA, SFDA, and FUDA methods by categorizing the baselines into three groups and simulating the OSUDA scenario for a fair comparison.



(a) Before adaptation (b) After adaptation

Figure 5. The t-SNE visualization on VisDA-2017. The red stars and solid dots represent the source and target samples, respectively (best viewed in color).

- **Direct** method conducts no adaptation, where the feature encoder F is pretrained on ImageNet and the cosine classifier C is directly initialized by source features.
- **UDA/FUDA** methods, including PCS [52], UCDIR [19], SDAT [35], CAF-A [47], CAF-D [47], ICON [53], ProtoOT [24] and MCUDA [25]. We simply modify the dataset interface to load the updated dataset and remove PCS components designed for unlabeled source data, as it's inaccessible in OSUDA.
- **SFDA** methods, including SHOT [27], NRC [49], AaD [50], SFDA-DE [9], BMD [34], GPUE [28], SFADA [15] and TPDS [39]. In the OSUDA setting, where pretrained source models are unavailable and training with only one sample per category is impractical, we initialize the backbone of all SFDA methods with a pretrained ResNet-50, as done in LCL. We then replace their original multi-layer classification head with a simplified single-layer cosine classifier, initialized using features from the one-shot source samples.

Evaluation Metrics. We use adaptation accuracy as our evaluation metric, as commonly used in domain adaptation methods (*e.g.*, PCS, UCDIR).

4.2. Results on Simulated OSUDA

On four public datasets, we randomly choose one labeled image per category in the source domain to simulate the OSUDA setting. To mitigate random effects, we conduct each experiment five times and report the mean results. Optimal results are in bold, and suboptimal results are underlined. Due to space constrains, only the mean results are presented in the main text; complete mean $\pm$ standard deviation results in Appendix B.2.

As shown in Tables 1-4, most existing UDA, FUDA, and SFDA methods show limited improvement over Direct, with some even performing worse. For instance, on VisDA-2017, six methods improve by less than 10% compared to Direct, with ProtoOT dropping by 5.82%. On Office-31, seven methods perform worse than Direct, with UCDIR showing a 20.58% decrease. On Office-Home, excluding PCS and ICON, SFDA-DE improves by only 3.62%, while SDAT decreases by 9.34%. On DomainNet, only UCDIR and GPUE exceed a 5% improvement. Performance inconsistency across datasets is also notable; for example, CAF-D outperforms Direct by 10.08% on VisDA-2017 but performs 2.15% worse on Office-Home. These results suggest that existing methods struggle in the OSUDA setting due to the limited source data, hindering both source domain knowledge acquisition and reliable cross-domain matching.

The above results show that the existing UDA, FUDA and SFDA methods cannot effectively improve the adaptation accuracy in the OSUDA setting. This is probably because the lack of source data makes it difficult for existing methods to either obtain source domain knowledge, or to eliminate the contingency of cross-domain matching results generated by one-shot sample.

In contrast, our LCL method, which learns discriminative features self-supervised and enforces stringent matching for cross-domain alignment significantly outperforms Direct, with gains of 23.97%, 13.26%, 16.8%, and 12.31% on VisDA-2017, Office-31, Office-Home, and DomainNet, respectively. These improvement are also 3.06%, 3.53%, 6.51%, and 6.39% higher than those of the best-performing comparative method, respectively.

Figure 5 shows t-SNE [41] plots of features before and after LCL adaptation on the VisDA-2017, demonstrating that LCL clusters samples effectively and aligns over half of the target clusters with their semantically consistent source samples after adaptation.

4.3. Results on Real-world OSUDA

To investigate DA methods in real-world OSUDA, we combine them with various face recognition backbones, including SphereFace[29], CosFace[43], ArcFace[6], VPL[8], AdaFace[23], and PartialFC[2], and compare their accuracy on the VSA dataset.

As shown in Table 6, the classification accuracies of all SOTA face recognition backbones are lower than 30%, highlighting the challenge of this real-world scenario. Nevertheless, LCL consistently gains significant improvement over Direct by 11.84% on average across six popular face recognition backbones, obviously higher than 7.39% of the best comparative method, PCS. These results validate LCL's effectiveness in real-world OSUDA and its robustness across different backbones. More experiments are elaborated in Appendix B.3.

Figure 6 shows some identification results of LCL on the VSA dataset, where we choose PartialFC as the backbone. In this challenging task, the domain gap between the source and target domain can be too large for LCL to work. Nevertheless, LCL can still successfully identify some blurry, occluded or angled faces in the surveillance scenario.

Table 5. Ablation study on Office-Home.

Method	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.
None (Direct)	18.25	30.42	36.65	24.72	27.02	29.11	31.13	22.81	44.76	35.58	23.00	41.90	30.45
$+\mathcal{L}_{\text{PC}}$	24.91	39.52	42.89	30.75	32.45	35.06	35.69	28.87	50.83	41.65	30.85	52.07	37.13
$+\mathcal{L}_{\text{ILC}}$	30.82	44.60	47.84	34.65	38.39	39.09	41.17	35.13	55.32	48.21	36.87	57.30	42.45
$+\mathcal{L}_{\text{CLC}}$ (w/o. BM)	30.44	47.30	52.78	35.43	40.95	43.26	45.56	34.89	59.07	50.78	37.38	59.95	44.82
$+\mathcal{L}_{\text{CLC}}$ (w. BM)	**32.55**	**49.57**	**55.91**	**38.88**	**46.18**	**46.95**	**48.17**	**36.43**	**61.03**	**51.77**	**37.58**	**61.99**	**47.25**

Table 6. Real-world OSUDA facial recognition performance (%) across various backbones on VSA.

Method	SphereFace	CosFace	ArcFace	VPL	AdaFace	PartialFC	Avg.
Direct	16.83	28.06	17.91	15.72	24.72	26.45	21.62
PCS	25.30	31.79	26.27	23.09	28.48	37.94	28.81
UCDIR	21.59	35.81	21.80	19.62	30.06	37.21	27.61
SDAT	2.90	2.40	2.40	2.30	4.90	2.30	2.87
CAF-A	3.84	3.42	4.29	4.47	3.31	1.58	3.49
CAF-D	3.28	3.53	4.41	3.59	4.37	1.78	3.49
ICON	3.58	3.02	3.08	2.71	2.87	2.92	3.03
ProtoOT	2.60	2.92	2.84	3.89	3.00	2.60	2.98
MCUDA	2.78	3.23	3.04	3.67	3.08	2.78	3.10
SHOT	9.89	20.14	11.75	13.07	8.65	19.72	13.87
NRC	13.91	22.11	14.80	14.12	19.77	19.59	17.38
SFDA-DE	14.49	26.45	14.25	14.33	26.19	25.59	20.22
BMD	13.78	22.22	15.07	14.20	22.14	21.61	18.17
GPUE	16.36	34.74	20.69	17.04	31.29	30.58	25.12
SFADA	11.41	19.12	13.23	14.46	6.23	16.91	14.30
TPDS	10.68	20.77	12.48	14.04	8.35	18.59	14.15
LCL	**28.79**	**44.31**	**31.16**	**25.48**	**42.78**	**44.70**	**36.20**

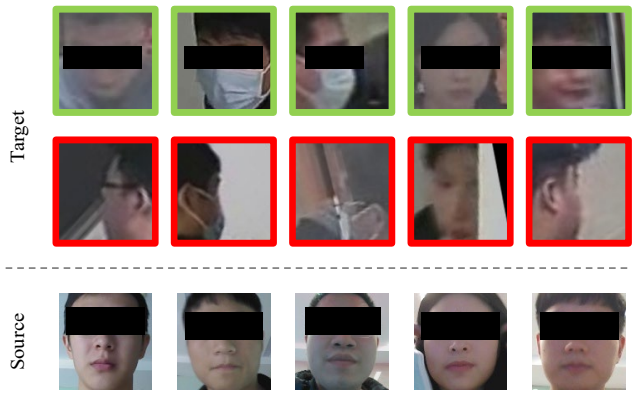


Figure 6. Examples of the successfully (green) and unsuccessfully (red) identified faces on the VSA dataset.

4.4. Ablation Study

In order to analyze the contribution of each component in the proposed LCL, we perform ablation study on the Office-Home dataset. Particularly, to analyze the contribution of our bidirectional matching (BM) scheme, we replace it with the simple matching scheme, which is denoted as $\mathcal{L}_{\text{CLC}}$ (w/o. BM) in Table 5. For a sample-prototype pair $(\boldsymbol{f}_i^s, \boldsymbol{\mu}_j^t)$, we set $r_{ij} = +1$ if C_j^t is the most similar cluster of $C_{P^s(i)}^s$, and $r_{ij} = -1$ otherwise.

As shown in Table 5, each component contributes to the final performance. Compared with the Direct baseline, $\mathcal{L}_{\text{PC}}$ improves the performance by 6.68%, demonstrating that semantic-aware discriminative feature learning is an essential component for domain adaptation. With the introduction of $\mathcal{L}_{\text{ILC}}$, $\mathcal{L}_{\text{PC}} + \mathcal{L}_{\text{ILC}}$ further gains an improvement of 5.32%, which demonstrates that $\mathcal{L}_{\text{ILC}}$ can further excavate the semantic information and improve the adaptation accuracy. And compared with $\mathcal{L}_{\text{PC}} + \mathcal{L}_{\text{ILC}}$, $\mathcal{L}_{\text{PC}} + \mathcal{L}_{\text{ILC}} + \mathcal{L}_{\text{CLC}}$ (w/o. BM) improves the average accuracy by 2.37%, showing the effectiveness of $\mathcal{L}_{\text{CLC}}$. Moreover, the introduction of bidirectional matching $\mathcal{L}_{\text{PC}} + \mathcal{L}_{\text{ILC}} + \mathcal{L}_{\text{CLC}}$ (w. BM) further improves the average accuracy by 4.80%, demonstrating that $\mathcal{L}_{\text{CLC}}$ facilitated by BM can more effectively align semantics across domains.

5. Conclusions and Future Work

This paper explores a realistic yet underexplored OSUDA problem, where each source category has only one sample, while the target domain has abundant unlabeled data. We propose an LCL method, which leverages in-domain contrastive learning to extract semantic-aware features and cross-domain contrastive learning to align semantically consistent clusters. Extensive experiments on four public benchmarks and a self-built real-world surveillance face recognition task validate LCL's effectiveness.

However, in some extreme cases, clustering errors may hinder reliable in-domain links, and the bidirectional matching may be too strict for cross-domain pairs (see Appendix C). Thus, our future work will extend binary hard links to probabilistic soft links to mitigate unreliable links' impact on the adaptation results.

6. Acknowledgments

This work is partially supported by National Natural Science Foundation of China under Grant (62372339, 62371350, 62072347), The numerical calculations in this paper have been done on the supercomputing system in the Supercomputing Center of Wuhan University.

References

[1] Xiang An, Xuhan Zhu, Yuan Gao, Yang Xiao, Yongle Zhao, Ziyong Feng, Lan Wu, Bin Qin, Ming Zhang, Debing Zhang, and Ying Fu. Partial FC: training 10 million identities on a single machine. In *IEEE International Conference on Computer Vision Workshops*, pages 1445–1449. IEEE, 2021. 6

[2] Xiang An, Jiankang Deng, Jia Guo, Ziyong Feng, Xuhan Zhu, Jing Yang, and Tongliang Liu. Killing two birds with one stone: Efficient and robust training of face recognition cnns by partial FC. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4032–4041. IEEE, 2022. 6, 7

[3] Lin Chen, Huaian Chen, Zhixiang Wei, Xin Jin, Xiao Tan, Yi Jin, and Enhong Chen. Reusing the task-specific classifier as a discriminator: Discriminator-free adversarial domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 7171–7180. IEEE, 2022. 2

[4] Gabriela Csurka, Riccardo Volpi, and Boris Chidlovskii. Unsupervised domain adaptation for semantic image segmentation: a comprehensive survey. *CoRR*, abs/2112.03241, 2021. 1

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. IEEE, 2009. 5

[6] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4690–4699. IEEE, 2019. 6, 7

[7] Jiankang Deng, Jia Guo, Debing Zhang, Yafeng Deng, Xiangju Lu, and Song Shi. Lightweight face recognition challenge. In *IEEE International Conference on Computer Vision Workshops*, pages 2638–2646. IEEE, 2019. 6

[8] Jiankang Deng, Jia Guo, Jing Yang, Alexandros Lattas, and Stefanos Zafeiriou. Variational prototype learning for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 11906–11915. IEEE, 2021. 6, 7

[9] Ning Ding, Yixing Xu, Yehui Tang, Chao Xu, Yunhe Wang, and Dacheng Tao. Source-free domain adaptation via distribution estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 7202–7212. IEEE, 2022. 3, 7

[10] Yaroslav Ganin and Victor S. Lempitsky. Unsupervised domain adaptation by backpropagation. In *Proceedings of the 32nd International Conference on Machine Learning*, pages 1180–1189. JMLR.org, 2015. 2

[11] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor S. Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 17:59:1–59:35, 2016. 1

[12] Rui Gong, Qin Wang, Dengxin Dai, and Luc Van Gool. One-shot domain adaptive and generalizable semantic segmentation with class-aware cross-domain transformers. *arXiv preprint arXiv:2212.07292*, 2022. 3

[13] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, pages 2672–2680, 2014. 2

[14] Arthur Gretton, Karsten M. Borgwardt, Malte J. Rasch, Bernhard Schölkopf, and Alexander J. Smola. A kernel two-sample test. *Journal of Machine Learning Research*, 13:723–773, 2012. 2

[15] Jiujun He, Liang Wu, Chaofan Tao, and Fengmao Lv. Source-free domain adaptation with unrestricted source hypothesis. *Pattern Recognition*, 149:110246, 2024. 7

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778. IEEE, 2016. 5

[17] Qichen He, Siying Xiao, Mao Ye, Xiatian Zhu, Ferrante Neri, and Dongde Hou. Independent feature decomposition and instance alignment for unsupervised domain adaptation. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 819–827. ijcai.org, 2023. 1

[18] Samitha Herath, Basura Fernando, Ehsan Abbasnejad, Munawar Hayat, Shahram Khadivi, Mehrtash Harandi, Hamid Rezatofighi, and Gholamreza Haffari. Energy-based self-training and normalization for unsupervised domain adaptation. In *IEEE International Conference on Computer Vision*, pages 11619–11628. IEEE, 2023. 1

[19] Conghui Hu and Gim Hee Lee. Feature representation learning for unsupervised cross-domain image retrieval. In *17th European Conference on Computer Vision*, pages 529–544. Springer, 2022. 2, 7

[20] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G. Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4893–4902. IEEE, 2019. 2

[21] Nazmul Karim, Niluthpol Chowdhury Mithun, Abhinav Rajvanshi, Han-Pang Chiu, Supun Samarasekera, and Nazanin Rahnavard. C-SFDA: A curriculum learning aided self-training framework for efficient source free domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 24120–24131. IEEE, 2023. 1

[22] Donghyun Kim, Kuniaki Saito, Tae-Hyun Oh, Bryan A. Plummer, Stan Sclaroff, and Kate Saenko. CDS: cross-domain self-supervised pre-training. In *IEEE International Conference on Computer Vision*, pages 9103–9112. IEEE, 2021. 1, 2, 4

[23] Minchul Kim, Anil K. Jain, and Xiaoming Liu. Adaface: Quality adaptive margin for face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 18729–18738. IEEE, 2022. 6, 7

[24] Bin Li, Ye Shi, Qian Yu, and Jingya Wang. Unsupervised cross-domain image retrieval via prototypical optimal transport. In *Thirty-Eighth AAAI Conference on Artificial Intelligence*, pages 3009–3017. AAAI Press, 2024. 7

[25] Jingzheng Li, Hailong Sun, Lei Chai, and Jiyi Li. Target structure learning framework for unsupervised multi-class

domain adaptation. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2024. 7

[26] Rui Li, Qianfen Jiao, Wenming Cao, Hau-San Wong, and Si Wu. Model adaptation: Unsupervised domain adaptation without source data. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 9638–9647. IEEE, 2020. 3

[27] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *Proceedings of the 37th International Conference on Machine Learning*, pages 6028–6039. PMLR, 2020. 2, 3, 7

[28] Mattia Litrico, Alessio Del Bue, and Pietro Morerio. Guiding pseudo-labels with uncertainty estimation for source-free unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 7640–7650. IEEE, 2023. 1, 2, 7

[29] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphereface: Deep hypersphere embedding for face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6738–6746. IEEE, 2017. 6, 7

[30] Yang Liu, Zhipeng Zhou, and Baigui Sun. COT: unsupervised domain adaptation with clustering and optimal transport. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 19998–20007. IEEE, 2023. 1

[31] Yawei Luo, Ping Liu, Tao Guan, Junqing Yu, and Yi Yang. Adversarial style mining for one-shot unsupervised domain adaptation. *Advances in neural information processing systems, 33:20612–20623*, 2020. 3

[32] Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko. Visda: The visual domain adaptation challenge. *arXiv preprint arXiv:1710.06924*, 2017. 5

[33] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *IEEE International Conference on Computer Vision*, pages 1406–1415. IEEE, 2019. 5

[34] Sanqing Qu, Guang Chen, Jing Zhang, Zhijun Li, Wei He, and Dacheng Tao. BMD: A general class-balanced multicentric dynamic prototype strategy for source-free domain adaptation. In *17th European Conference on Computer Vision*, pages 165–182. Springer, 2022. 7

[35] Harsh Rangwani, Sumukh K. Aithal, Mayank Mishra, Arihant Jain, and Venkatesh Babu Radhakrishnan. A closer look at smoothness in domain adversarial training. In *Proceedings of the 39th International Conference on Machine Learning*, pages 18378–18399. PMLR, 2022. 2, 7

[36] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *11th European Conference on Computer Vision*, pages 213–226. Springer, 2010. 5

[37] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, Trevor Darrell, and Kate Saenko. Semi-supervised domain adaptation via minimax entropy. In *IEEE International Conference on Computer Vision*, pages 8049–8057. IEEE, 2019. 2

[38] Hui Tang and Kui Jia. Discriminative adversarial domain adaptation. In *Thirty-Fourth AAAI Conference on Artificial Intelligence*, pages 5940–5947. AAAI Press, 2020. 2

[39] Song Tang, An Chang, Fabian Zhang, Xiatian Zhu, Mao Ye, and Changshui Zhang. Source-free domain adaptation via target prediction distribution searching. *International journal of computer vision, 132(3):654– 672*, 2024. 7

[40] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474*, 2014. 2

[41] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9 (11), 2008. 7

[42] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5385–5394. IEEE, 2017. 5

[43] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 5265–5274. IEEE, 2018. 6, 7

[44] Xinyi Wu, Zhenyao Wu, Yuhang Lu, Lili Ju, and Song Wang. Style mixing and patchwise prototypical matching for one-shot unsupervised domain adaptive semantic segmentation. *In Proceedings of the AAAI Conference on Artificial Intelligence, pages 2740–2749*, 2022. 3

[45] Haifeng Xia, Handong Zhao, and Zhengming Ding. Adaptive adversarial network for source-free domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 8990–8999. IEEE, 2021. 3

[46] Haifeng Xia, Taotao Jing, and Zhengming Ding. Maximum structural generation discrepancy for unsupervised domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3434–3445, 2023. 2

[47] Binhui Xie, Shuang Li, Fangrui Lv, Chi Harold Liu, Guoren Wang, and Dapeng Oliver Wu. A collaborative alignment framework of transferable knowledge extraction for unsupervised domain adaptation. *IEEE Transactions on Knowledge and Data Engineering*, 35(7):6518–6533, 2023. 1, 7

[48] Shiqi Yang, Yaxing Wang, Joost van de Weijer, Luis Herranz, and Shangling Jui. Generalized source-free domain adaptation. In *IEEE International Conference on Computer Vision*, pages 8958–8967. IEEE, 2021. 2

[49] Shiqi Yang, Yaxing Wang, Joost van de Weijer, Luis Herranz, and Shangling Jui. Exploiting the intrinsic neighborhood structure for source-free domain adaptation. In *Advances in Neural Information Processing Systems*, pages 29393–29405, 2021. 7

[50] Shiqi Yang, Yaxing Wang, Kai Wang, Shangling Jui, and Joost van de Weijer. Attracting and dispersing: A simple approach for source-free domain adaptation. In *Advances in Neural Information Processing Systems*, pages 5802–5815, 2022. 2, 4, 7

[51] Shiqi Yang, Yaxing Wang, Joost van de Weijer, Luis Herranz, Shangling Jui, and Jian Yang. Trust your good friends:

Source-free domain adaptation by reciprocal neighborhood clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12):15883–15895, 2023. 1

[52] Xiangyu Yue, Zangwei Zheng, Shanghang Zhang, Yang Gao, Trevor Darrell, Kurt Keutzer, and Alberto L. Sangiovanni-Vincentelli. Prototypical cross-domain self-supervised learning for few-shot unsupervised domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 13834–13844. IEEE, 2021. 1, 2, 3, 4, 5, 6, 7

[53] Zhongqi Yue, Qianru Sun, and Hanwang Zhang. Make the u in uda matter: Invariant consistency learning for unsupervised domain adaptation. *Advances in Neural Information Processing Systems,36:26991–27004*, 2023. 7

[54] Qianyu Zhou, Qiqi Gu, Jiangmiao Pang, Xuequan Lu, and Lizhuang Ma. Self-adversarial disentangling for specific domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7):8954–8968, 2023. 2

[55] Wanqing Zhu, Jia-Li Yin, Bo-Hao Chen, and Ximeng Liu. Srouda: Meta self-training for robust unsupervised domain adaptation. In *Thirty-Seventh AAAI Conference on Artificial Intelligence*, pages 3852–3860. AAAI Press, 2023. 1