Unveiling the Unknown: Unleashing the Power of Unknown to Known in Open-Set Source-Free Domain Adaptation

Fuli Wan*, Han Zhao*, Xu Yang†, and Cheng Deng†
School of Electronic Engineering, Xidian University, Xi'an710071, China
{wfl128.xidian, hzhao1698, xuyang.xd, chdeng.xd}@gmail.com

Abstract

Open-Set Source-Free Domain Adaptation aims to transfer knowledge in realistic scenarios where the target domain has additional unknown classes compared to the limited-access source domain. Due to the absence of information on unknown classes, existing methods mainly transfer knowledge of known classes while roughly grouping unknown classes as one, attenuating the knowledge transfer and generalization. In contrast, this paper advocates that exploring unknown classes can better identify known ones, and proposes a domain adaptation model to transfer knowledge on known and unknown classes jointly. Specifically, given a source pre-trained model, we first introduce an unknown diffuser that can determine whether classes in space need to be split and merged through similarity measures, to estimate and generate a wider class space distribution, including known and unknown classes. Based on such a wider space distribution, we enhance the reliability of known class knowledge in the source pre-trained model through contrastive constraint. Finally, various supervision information, including reliable known class knowledge and clustered pseudo-labels, optimize the model for impressive knowledge transfer and generalization. Extensive experiments show that our network can achieve superior exploration and knowledge generalization on unknown classes, while with excellent known class transfer. The code is available at https://github.com/xdwfl/UPUK.

1. Introduction

By transferring knowledge from label-rich source domains to unlabeled target domains, unsupervised domain adaptation (UDA) has exhibited huge potential to address the challenges of costly data labeling and domain distribution shifts in real-world scenarios. Existing methods mainly learn the domain invariant representation by distribution matching [19, 20, 34, 40, 46] and adversarial optimization [7, 21, 35], to mitigate domain distribution shift. They can achieve impressive transfer effects by following the close-set and real-time data-accessible assumption, which shares the same class space in source and target domains, and accesses the source data at any time. However, in the real world, the open-set phenomenon, where the target domain has additional unknown class space to the source domain, always appears along with limited access to source data [5, 16]. In such most realistic scenarios, these methods fail to transfer knowledge without accessing source data and shared class distribution, making the knowledge transfer in such scenarios challenging.

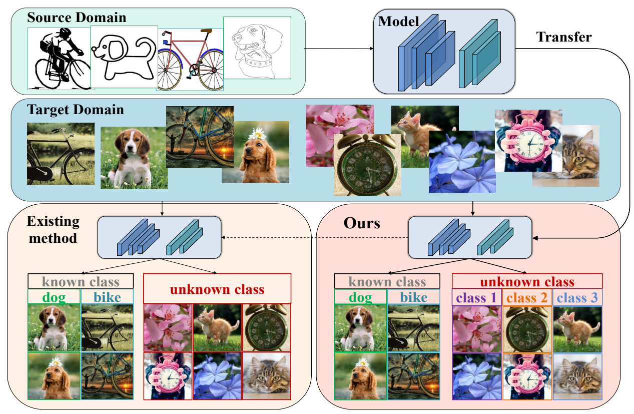


Figure 1. Comparison of existing OS-SFDA methods and our proposed method. Existing methods ignore the unknown classes by grouping unknown classes as one, while our method can further explore the unknown class information.

Recently, to tackle such a challenging and realistic task, the open-set source-free domain adaptation (OS-SFDA) has emerged to perform knowledge transfer in the open-set scenario by accessing only the source pre-trained model, not the source data. Considering the lack of information on target unknown classes, existing approaches [6, 22, 31, 37] focus on transferring knowledge of known classes through pseudo-supervision, while grouping unknown classes as

*Equal contribution.
†The corresponding authors.

one, as seen Fig. 1. However, such a grouping way restricts the thorough exploration of target class information, thereby diminishing models' transfer and generalization abilities on known and unknown classes, respectively.

To this end, we present a domain adaptation network for OS-SFDA, to explore target class space and transfer knowledge on known and unknown classes jointly. Specifically, given the source pre-trained model, an unknown diffuser is introduced to determine to-be-split/merged classes through similarity measures, generating a wider class space distribution, including known and unknown classes. Based on a wider distribution in the feature space, the reliability of the known class knowledge from the source pre-trained model is enhanced through contrastive constraint. Finally, the reliable known class knowledge, along with the clustered pseudo-labels spanning the extensive class space, can collaboratively serve as supervision during model optimization. This collaborative approach results in notable knowledge transfer and enhanced generalization capabilities. Our contributions are summarized as follows:

- We propose a domain adaptation network that is the first to excavate the unknown class space for boosting knowledge transfer and generalization to the target domain in OS-SFDA.
- We design an unknown diffuser to explore wider and more precise target class space in the target domain, further benefiting known knowledge transfer and unknown generalization.
- To realize effective known knowledge transfer and unknown generalization, we utilize the reliable known knowledge and clustered pseudo-labels in the broader class space, to serve as supervision during optimization.
- Experiments can demonstrate that our method can explore target class space precisely, and yield state-of-the-art results among multiple OS-SFDA benchmarks.

2. Related Work

2.1. Source-Free Domain Adaptation

In recent years, with accessing only the source pre-trained model for transferring knowledge to the target domain, source-free domain adaptation (SFDA) methods [4, 16, 27, 41, 43–45, 47]have been widely used for overcoming source data privacy to achieve knowledge transfer in real-world scenarios. The existing methods are mainly divided into two categories, including methods based on pseudo-labeling [14, 16] and based on neighborhood clustering [43, 44]. For the former approaches, the classical method – SHOT [16] freezes the source classifier and optimizes the feature extractor, through mutual information maximization and source-target feature alignment based on self-supervised pseudo-labeling. To fully utilize the source and target domain knowledge, CoWA-JWDS [14] introduces a new sample confidence score – JMDS better to exert the importance of pseudo-label confidence in distinguishing samples. For the latter approaches, G-SFDA [43] proposes a method of clustering target features and semantically similar neighborhoods to make the model better adapt to the target domain. AaD [44] has introduced prediction consistency of local neighborhood features and discrete prediction of different potential features, which achieve efficient feature clustering and distribution to better transfer knowledge. In addition, there are also some approaches to tackling SFDA from other aspects. U-SFAN [27] uses predicted quantified uncertainty to guide target adaptation. CRS [47] embeds class relationship similarity to transfer domain-invariant class relationships.

2.2. Open-Set Domain Adaptation

Recently, open-set domain adaptation (OSDA) tackles the realistic open-set challenge, where the target domain has additional unknown class space to the source domain. Currently, existing methods [2, 10, 15, 18, 29, 30, 33, 38] transfer known class knowledge while grouping unknown classes as one. For example, OSBP [30] uses pseudo labels to guide classifier learning and to construct constraints for the optimization of the model. Later, to better utilize pseudo-labels, STA [18] proposes an end-to-end method to gradually separate known and unknown classes while aligning feature distributions. OVANET [29] introduces the one-vs-all classifier to learn the inter- and intra-class distance, and minimizes class entropy to distinguish known and unknown classes. OSLPP [38] learns a common subspace from the source and target domains while gradually selecting and rejecting pseudo-labeled target data to promote the model's transfer on known classes. By introducing adversarial learning, UADAL [10] aligns the source- and the target-known distribution while segregating the target-unknown distribution, achieving better knowledge transfer.

2.3. Open-Set Source-Free Domain Adaptation

When the open-set phenomenon appears, along with limited access to source data, the methods without accessing source data and shared class distribution cannot transfer favorable knowledge, causing them to fail in real-world scenarios. OS-SFDA [6, 11, 13, 17, 22, 26, 37, 42] has emerged to perform knowledge transfer in the open-set scenario with accessing only the source pre-trained model, not the source data. Specifically, the existing methods generally transfer known knowledge and group unknown classes as one. Among these methods, without source data, FS [13] designs the self-adaptive model to transfer the source task-specific knowledge to the target domain effectively, and the target unknown classes are grouped into one class under the open-set setting due to the lack of unknown class information. To tackle OS-SFDA, OSHT [6] adopts pseudo-labeling for adaptation and the entropy-based metric to reject unknown classes. Distill-SODA [37] introduces novel style-based adversarial data augmentation and the closed-set affinity score for better knowledge distillation from a self-supervised ViT during adaptation. SF-PGL [22] introduces a balanced pseudo-labeling strategy to progressive graph learning for adaptation and distinguishes known and unknown classes by applying the confidence threshold. In addition, the SFDA methods [14, 16, 44, 47] separate known and unknown classes and apply their strategies mentioned in Sec. 2.1 for known knowledge transfer under the open-set setting to tackle the OS-SFDA.

However, under the OS-SFDA setting, where the target domain has a additional abundant class distribution than the source domain, mentioned OS-SFDA methods group unknown classes as one, which restricts the thorough exploration of target class information, hampering knowledge transfer and generalization. Our work advocates exploring unknown classes can better identify known ones. It should be noted that recently some methods have attempted to explore unknown classes in OSDA [11, 23, 49], which requires source data access to realize domain adaptation. On the contrary, our focused OS-SFDA needs no source data for exploring unknown class space that is more realistic and challenging, essentially differing from them. Concretely, their source data access results in the failure to deal with the most realistic OS-SFDA. Moreover, they utilize 'unknown class number prior' to predefine [11, 23] or enumerate [49] unknown class numbers to explore the unknown class space, whose reliance on unknown prior further limits their scalability in real world. Instead, during unknown class exploration, our method requires no source data and unknown class number prior, truly achieving spontaneous and applicable unknown class exploration in OS-SFDA. We present a domain adaptation network for OS-SFDA, which can fully explore the target class space to transfer knowledge on known classes and generalize on unknown classes.

3. Method

In this section, we will introduce the details of our proposed method, and the framework is depicted in Fig. 2.

Problem Setting. OS-SFDA aims to solve the knowledge-transferring problem in more realistic scenarios, where the source data have restricted access, and the target domain exhibits wider class distribution than the source domain. In such a task with restricted-accessible source data, we are given an unlabeled target domain $\mathcal{D}^t = \{(\boldsymbol{x}_i^t)\}_{i=1}^n$ with n target samples, and a pre-trained source model. More importantly, the class space $\mathcal{Y}^t$ of the target domain is broader than the one $\mathcal{Y}^s$ of the source domain: $\mathcal{Y}^t = \mathcal{Y}^s \cup \mathcal{Y}^{ukn}$. In other words, in addition to the source domain's class space $\mathcal{Y}^s$, $\mathcal{Y}^t$ also contains an additional unknown class space $\mathcal{Y}^{ukn}$. Here we call $\mathcal{Y}^s$ and $\mathcal{Y}^{ukn}$ as the target domain's known class and unknown class spaces, respectively, whose numbers of categories are denoted as K_n and K_{un}. The source domain pre-trained model is obtained by pre-training on the source domain in $\mathcal{Y}^s$, mainly consisting of feature extractor $f(\cdot)$ and classifier $g(\cdot)$. Therefore, it is expected to enable the source pre-trained model to overcome the class distribution discrepancy between domains, while achieving further knowledge transfer on the known classes and exploration on the unknown classes.

3.1. Exploration of the Target Class Space

Inspired by the effectiveness of no-parameter clustering methods [1, 25, 32, 48] in exploring the class space, we introduce an unknown diffuser based on the Multilayer Perceptron (MLP), which includes the main network C and the sub-network C_k^{sub} in the current class space, whose goal is to explore the target class space by inferring the target class number K. Reasonably, we first initialize the target class space with the source known class space; that is, we assume that the initial value of the target class number K is the number of known classes K_n. The unknown diffuser progressively explores a wider class space in the target domain through continuous inferring and optimization of K. Note that the value of K changes dynamically with the inference process during training. The unknown diffuser completes the exploration of the target class space by performing two training processes: optimization in the latest class space and hierarchical class exploration.

Optimization in Latest Class Space. Since the target domain has no labels during exploration, we need to use the clusters obtained by clustering the target data to guide us in exploring the target class space. Simultaneously, we transform the exploration of the target class space into the exploration of the target cluster number. In this part, we learn the discriminative cluster assignments through the constraints of the feature-cluster assignment, which is beneficial for exploring the target cluster number. Firstly, the feature is the output of feature extractor $f(\cdot)$ in the source pre-trained model, denoted as $\boldsymbol{z}_i = f(\boldsymbol{x}_i^t) \in \mathbb{R}^d$, where d is the dimension of the feature space, i represents the i-th sample. The features of n samples are denoted as $\boldsymbol{Z}$. Then, the main network C in the unknown diffuser maps the feature into the cluster assignment $\widetilde{\boldsymbol{Z}} = c(\boldsymbol{Z})$, where $c(\cdot)$ is the mapping operation of C, and the cluster assignment of feature $\boldsymbol{z}_i$ is denoted as $\widetilde{z}_i \in \mathbb{R}^K$. The discriminative cluster assignments under the current class space (K) are obtained by optimizing the cluster assignment under the guidance of the discriminative feature $\boldsymbol{Z}$. Specifically, to obtain the pseudo cluster-assignment $Y^{\boldsymbol{Z}} = \{y_i^{\boldsymbol{Z}}\}_{i=1}^n, y_i^{\boldsymbol{Z}} \in \mathbb{R}$, we first use k-means under current target class space (K) on the feature set $\boldsymbol{Z}$. Then, according to $Y^{\boldsymbol{Z}}$, a teacher cluster soft-distribution $\widetilde{\boldsymbol{Y}} \in \mathbb{R}^{n \times K}$ is generated based on the feature:

$$\widetilde{y}_{i,k} = \|\boldsymbol{z}_i - \boldsymbol{\eta}_k\|_{\ell_2}^2 , \tag{1}$$

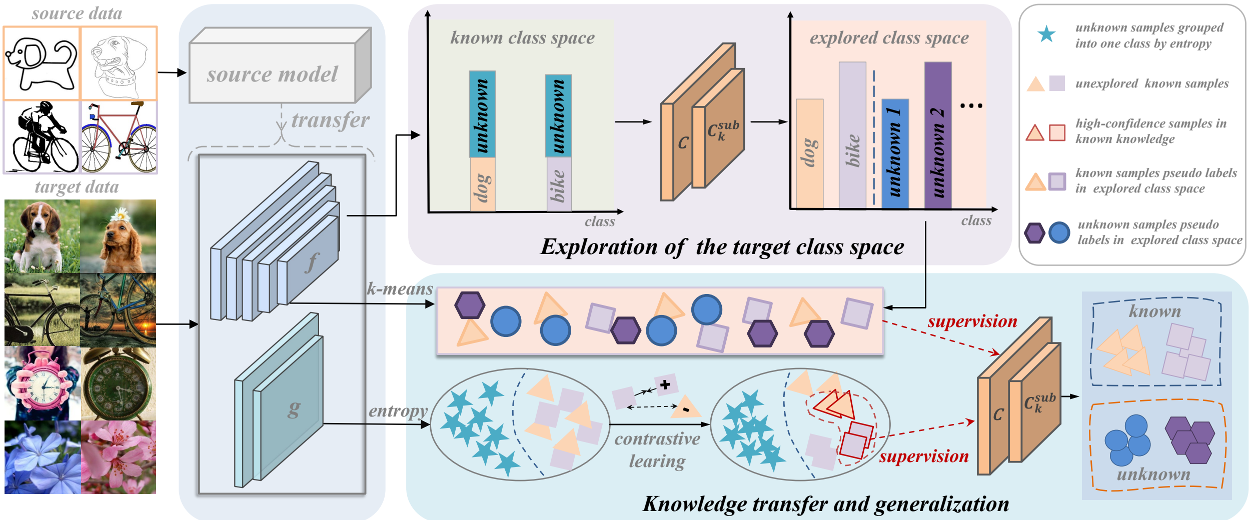


Figure 2. The framework of our proposed method. Given the source pre-trained model, we first introduce an unknown diffuser on its initial known class space. The unknown diffuser can determine whether classes in space need to be split and merged through similarity measures, to estimate and generate a wider class space distribution, including known and unknown classes. Based on such a wider class space, we improve the reliability of the known class knowledge inside the source model by contrastive learning. In this way, the reliable known class knowledge and clustered pseudo-labels over the wider class space, can jointly serve as supervision during model optimization, leading to powerful knowledge transfer and generalization capabilities.

where $\widetilde{y}_{i,k}$ is the k-th element in $\widetilde{\boldsymbol{y}}_i$, and $\widetilde{\boldsymbol{y}}_i$ is the teacher cluster soft-distribution of feature $\boldsymbol{z}_i$. $\boldsymbol{\eta}_k = \sum_{i+1}^{n_k} \boldsymbol{z}_i^k / n_k$ is the center feature of the k-th center and n_k is the feature number of the k-th master cluster. And $\boldsymbol{z}_i^k$ is the i-th element in $\boldsymbol{Z}^k \in \mathbb{R}^{n_k \times d}$, which is the feature matrix of those features with $y_i^{\boldsymbol{Z}} = k$. Note that the teacher cluster soft-distribution of each feature is normalized. Then, we utilize the alignment loss to enforce the consistency of teacher cluster soft-distribution $\widetilde{\boldsymbol{Y}}$ with the cluster assignment $\widetilde{\boldsymbol{Z}}$:

$$\mathcal{L}_{alg} = \sum_{i=1}^{n} L_{alg}(x_i^t) = \sum_{i=1}^{n} \widetilde{\boldsymbol{z}}_i \cdot \widetilde{\boldsymbol{y}}_i. \tag{2}$$

The cluster assignment under the current target class space (K) can be promoted through optimization on $c(\cdot)$ in the training process, so as to promote the exploration of the target class space.

Hierarchical Class Exploration. To fully explore the target class space by inferring the precise target class number K, inspired by [3], the Metropolis-Hastings framework (M-H) is utilized as the underlying framework in the class diffusion process, which computes the Hasting ratio for changing K through splitting and fusion. Besides that, inspired by the hierarchical clustering [39], we design a hierarchical class exploration method to explore the target class space based on the M-H framework, that performs splitting and merging on more dispersive and coupled clusters, respectively. Firstly, we obtain the discriminative cluster assignment after splitting/merging in all possible splitting/merging cases. We then select the potential clusters to be split or merged based on the assignment distance between clusters. In the end, these selected clusters are put into the candidate set to be split or merged, while we use the M-H framework [8] on those clusters in the candidate set to determine the splitting and merging decisions. We assume that the current cluster distributions and those under all possible splitting/merging cases are called the distributions before and after splitting/merging.

Specifically, we first obtain the discriminative cluster distribution after splitting/merging in all possible splitting/merging cases, in which the cluster before splitting/merging is already learned above. Then, the cluster distribution after merging can be acquired directly by merging the assignments of two clusters. However, obtaining the sub-cluster distribution after splitting is apparently unreasonable, by randomly dividing the target data in the master cluster. For subsequent convenience, we define the clusters before/after splitting as the master cluster and sub-cluster, respectively. To gain the discriminative sub-cluster distribution, we introduce the sub-network C_k^{sub} into the unknown diffuser, and C_k^{sub} contains K sub-clustering layers with the output dimension equaling to 2. One of these sub-network C_k^{sub} is applied on each master cluster to generate the sub-cluster assignment $\widetilde{\boldsymbol{Z}}^{sub,k} = c_k^{sub}(\boldsymbol{Z}^k)$, in which $c_k^{sub}(\cdot)$ represents the operation of the k-th sub-network applying on the k-th master cluster. Then, to obtain the pseudo-sub-cluster assignment, the k-means first is performed on the feature matrix $\boldsymbol{Z}^k$ of the k-th master cluster, similar to cluster optimization with Eq. (1), Subsequently, a teacher sub-cluster assignment $\widetilde{\boldsymbol{Y}}^k \in \mathbb{R}^{n_k \times 2}$ is also generated for the k-th master cluster:

$$\widetilde{y}^k_{i,j} = \left\| \boldsymbol{z}^k_i - \boldsymbol{\eta}^k_j \right\|^2_{\ell_2}, \tag{3}$$

where $\widetilde{y}^k_{i,j}$ is the i-th row and j-th column element in $\widetilde{\boldsymbol{Y}}^k$. $\boldsymbol{\eta}^k_j$ is the center feature of the j-th sub-cluster in k-th master cluster. And we utilize the alignment loss to optimize them:

$$\mathcal{L}^{\text{sub}}_{alg} = \sum_{k=1}^{K} \sum_{i=1}^{n_k} \widetilde{\boldsymbol{z}}^{sub,k}_i \cdot \widetilde{\boldsymbol{y}}^k_i, \tag{4}$$

where $\widetilde{\boldsymbol{z}}^{sub,k}_i$ is the i-th element in $\widetilde{\boldsymbol{Z}}^{sub,k}$. The training process on the sub-network C^{sub}_k is synchronous with the one on the main network C.

In addition, hierarchical clustering [39] gradually divides data by merging pairs of clusters with smaller distances or splitting the two most distant sub-clusters in one cluster. According to the assignment distance between clusters, the candidate sets to be split/merged are generated, inspired by the hierarchical clustering. In our method, to explore more different classes, those cluster pairs (sub-cluster pairs) with far assignment distances are considered as those likely to be split. While those with close assignment distances are considered as those likely to be merged due to high similarity between two main clusters. The candidate sets to be split/merged based on this can be written as:

$$CS^{ED}_{merg} = \{(k_1, k_2)| \in \arg\text{sort}_{n_{if}} ED(\boldsymbol{\eta}_{k_1}, \boldsymbol{\eta}_{k_2})\}, \tag{5}$$

$$CS^{ED}_{spl} = \{k| \in \arg\text{sort}_{n_{if}} - ED(\boldsymbol{\eta}^k_1, \boldsymbol{\eta}^k_2)\}, \tag{6}$$

where $ED(\cdot,\cdot)$ computes the Euclidean distance of the cluster embedding center. n_{if} are the number of cluster(s) with potential to be merged and split.

After obtaining the candidate set, we can infer more accurate K and explore a more accurate target class space, by calculating the Hasting ratio from the M-H framework [8] among (sub)cluster pairs in the candidate set and comparing it with 1. More details are provided in the supplementary material. In this way, we use the inferred latest target class number K to update the unknown diffuser (the output dimension of C and the number of sub-clustering layers) at the next optimization step.

During the optimization in the latest more accurate class space, we learn the discriminative cluster distribution under the latest updated target class number K. According to these cluster distributions, we decide the splitting and merging decisions to change the target class number K in the process of hierarchical class exploration. The more precise K and corresponding cluster assignments are generated, through operating these two processes iteratively, which is beneficial to fully explore the target class space.

Algorithm 1 The training procedure of our proposed model

Input: data, source pre-trained model, main network C and sub-network C^{sub}_k of unknown diffsuer

// Exploration of the target class space
for $iter_1 = 1$ to $iter^{max}_1$ **do**
  Obtain clustered pesudo-distribution $\widetilde{\boldsymbol{Y}}$ and $\widetilde{\boldsymbol{Y}}^{\boldsymbol{k}}$ by k-means
  Optimize C and C^{sub}_k guided by $\widetilde{\boldsymbol{Y}}$ and $\widetilde{\boldsymbol{Y}}^{\boldsymbol{k}}$ in Eqs. (2) and (4)
  Determine CS_{slp} / CS_{merg} by similarity measurement in Eqs. (5) and (6)
  Diffuse the class space by updating K on CS_{slp} / CS_{merg}, based on M-H framework
  Update the unknown diffuser by the updated K
end for
// Knowledge transfer and generalization
for $iter_2 = 1$ to $iter^{max}_2$ **do**
  Promote reliability of known knowledge by contrastive constraint in Eq. (7)
  Jointly optimize C and the pre-trained model in the explored class space via $\mathcal{L}_{kt}$ and $\mathcal{L}_{alg}$ in Eq. (8)
end for

3.2. Knowledge Transfer and Generalization

After exploring the target class space in Sec. 3.1, we aim to transfer knowledge on known classes more effectively and improve the model's generalization on unknown classes. In this subsection, we promote the reliability of known knowledge and further optimize over wider class distribution.

Promotion of Known Knowledge Reliability. To better utilize known knowledge, we hope to improve the reliability of known knowledge obtained by the classification layer of the pre-trained source model during training. The input of the model is all target samples, including known and unknown samples, from which we need to select the known ones. According to the logit output of known samples having smaller entropy than the unknown ones, we select the top 50% of samples with the lowest entropy as known samples, which are placed in set $\mathcal{M}$. After selection, the contrastive constraint $\mathcal{L}_{con}$ is utilized on $\mathcal{M}$ to enhance the discrimination of the known samples. Concretely, it brings the known features of similar predictions closer and those of dissimilar predictions farther away, achieving discrimination feature assignment to strengthen the reliability of known knowledge. The contrastive constraint is written as:

$$\mathcal{L}_{con} = -\sum_{i \in \mathcal{M}} \boldsymbol{p}^T_i \cdot \boldsymbol{p}^+_i + \sum_{i \in \mathcal{M}} \boldsymbol{p}^T_i \cdot \boldsymbol{p}^-_i, \tag{7}$$

where $\boldsymbol{p}_i = \sigma(g(\boldsymbol{z}_i)) \in \mathbb{R}^d$) is the output of classifier g and σ is the softmax function.

Optimization over Wider Class Distribution. After that, we optimize the unknown diffuser over the wider class space by taking advantage of the reliable known knowledge and alignment constraint on the explored space. In such an optimization based on the explored wider class space, we first select the high-confidence known samples as the reliable known knowledge. Then, we hope to use the hard pseudo-labels of the high-confidence known samples as supervision information, thereby leveraging them for better knowledge transfer on the explored class space. Specifically, for a known sample, we take the maximum logit output by the classifier in the source pre-trained model as its confidence. And the known sample is considered as the high-confidence known sample when it is greater than a manually-set confidence threshold, whose hard pseudo-label is represented by the one-hot paradigm $\boldsymbol{y}_i^h$ ($x_i^h \in \mathcal{H}$). Note that $\mathcal{H}$ is the set of the high-confidence known samples. Therefore, we can perform better knowledge transfer on the explored class space by leveraging the reliable known knowledge $\boldsymbol{y}_i^h$ obtained on the source pre-trained model. In addition, we also perform alignment constraint on the explored class space, similar to Eq. (2), to complete the transfer on known samples further and achieve generalization on unknown samples. Finally, we leverage the supervision of the high-confidence known sample labels and the alignment constraint, to jointly achieve superior knowledge transfer on known classes and generalization on unknown classes over the wider class space:

$$\begin{aligned}\mathcal{L} = \mathcal{L}_{kt} + \mathcal{L}_{alg} &= \sum_{i}^{\mathcal{H}} L_{kt}(\widetilde{\boldsymbol{z}}_i, \boldsymbol{y}_i^h) + \sum_{i=1}^{n} L_{alg}(x_i^t) \\ &= \underbrace{\sum_{i}^{\mathcal{H}} L_{kt}(\widetilde{\boldsymbol{z}}_i, \boldsymbol{y}_i^h) + \sum_{j_1=1}^{n_1} L_{alg}(x_{j_1}^t)}_{known} + \underbrace{\sum_{j_2=1}^{n_2} L_{alg}(x_{j_2}^t)}_{unknown},\end{aligned} \tag{8}$$

where $\mathcal{L}_{kt}$ represents the guidance of high-confidence known samples, which is the cross-entropy loss between their pseudo-labels and soft distributions. $\mathcal{L}_{alg}$ is the alignment constraint to achieve superior knowledge transfer on known classes and generalization on unknown classes. $\widetilde{\boldsymbol{z}}_i$ represents the soft-distribution output by the main network C on the explored class space. n_1 and n_2 represent the number of known and unknown samples, respectively.

4. Experiments

4.1. Experimental Setup

Datasets. We evaluate our method on three benchmark datasets for OS-SFDA through experiments, which are **Office-31** [28], **Office-Home** [36], and **VisDA-C 2017** [24]. Due to limited space, the results and analysis on **VisDA-C 2017** are summarized in supplementary material.

- **Office-31** [28] consists of three domains (Amazon, Webcam, DSLR), with 31 classes and 4652 images of common objects. Following the protocols [18] on Office-31 and Office-Home, we select the first 20 classes as the shared classes of the source and target domain while the remaining 11 classes are considered unknown.
- **Office-Home** [36] is a more challenging dataset. It contains four domains (Art, Product, Clipart, RealWorld) with 65 classes and 15500 images of common objects. Like Office-31, we select the first 25 classes as classes shared by the source and target domain, and the remaining 40 classes are considered unknown.

Baselines. For OS-SFDA, we compare our method with the clustering on Resnet [9], the classic method SHOT [16], and the advanced method AaD [44], which do sufficient experiments under the OS-SFDA setting. In addition, our method uses the unknown diffuser based on clustering (class number-agnostic) to explore the target class space. For a fair comparison, we thoroughly compare our method with the no-parameter clustering methods (FINCH [32], COMIC [25], DenMune [1]).

Evaluation. As shown in Tab. 1 and Tab. 2, we report domain adaptation results for OS^* and UNK, which represent the average accuracy of the known and the unknown class, respectively. Simultaneously, $HOS = \frac{2 \times OS^* \times UNK}{OS^* + UNK}$ is denoted as the harmonic mean between OS^* and UNK, which is the key indicator for evaluating performance. Moreover, the inferred target class number K and accuracy ACC are used to evaluate the effectiveness of exploring the target class space in Tab. 3 and Tab. 4. The setting of ACC follows the previous protocols [12].

Implementation Details. For the source pre-trained model, we utilize the backbone of ResNet-50 [9] for Office-31 and Office-Home. For a fair comparison, we use the same network architecture with SHOT [16] and AaD [44], which includes a backbone and two additional fully connected layers. We adopt SGD with a momentum of 0.9 and a batch size of 64 for all datasets. The learning rate is set to 1e-3 on all datasets for all layers of the backbone, and the last two added fully-connected layers are applied 1e-2. We train 15 epochs for all datasets. The unknown diffuser includes one convolutional layer with 256 hidden dimensions. For a fair comparison, our network and the compared no-parameter clustering methods use the same input features, output by the source pre-trained model.

4.2. Results and Analysis

As shown in Tab. 1 and Tab. 2, the UNK and HOS for our method are much higher than SHOT on all datasets. Our method achieves 10.4% and 2.3% relative improvements over the state-of-art AaD for average OS* and HOS on Office-31, respectively. On the relatively small-scale Office-31 with fewer classes, our method demonstrates competitiveness across various indicators, standing in con-

method	SF	A→D			A→W			D→A			D→W			W→A			W→D			Avg		
		OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS
Cluster	✗	10.0	64.0	17.0	7.0	61.0	13.0	5.0	69.0	10.0	8.0	49.0	14.0	5.0	64.0	10.0	10.0	47.0	17.0	7.5	59.0	13.5
SHOT	✓	87.2	50.0	55.7	89.0	38.6	53.8	78.2	44.2	56.5	97.4	72.3	83.0	73.5	47.6	57.8	89.6	75.0	81.7	**85.8**	54.6	64.8
AaD	✓	69.9	86.7	77.4	62.0	83.9	71.3	60.8	81.8	69.8	83.2	97.4	89.8	58.4	81.5	68.0	82.3	95.2	88.3	69.4	**87.8**	77.4
Ours(w/o s)	✓	78.4	73.4	75.8	87.9	90.3	89.1	72.3	76.1	74.1	92.8	80.5	86.2	70.9	76.8	73.7	85.8	79.8	82.7	<u>81.4</u>	79.5	**80.3**
Ours(w/ s)	✓	82.3	71.8	76.7	90.0	72.0	80.0	72.0	79.4	75.5	79.5	87.3	83.2	74.2	82.7	78.2	80.6	88.3	84.3	79.8	<u>80.3</u>	<u>79.7</u>

Table 1. Accuracy (%) on Office-31 for OS-SFDA. The three indicators(OS*, UNK, HOS) have been explained in Sec. 4.1. 'w/o s' and 'w/ s' indicate with or without supervision using reliable known knowledge. Note that 'A→D' represents that the model per-trained on the source domain Amazon transfers to the target domain DSLR. Bold and ' _ ' represent the best and second results, respectively.

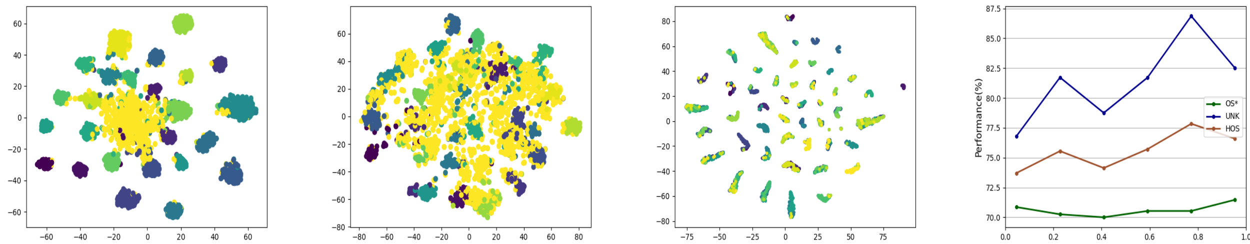
method	SF	Ar→Cl			Ar→Pr			Ar→Rw			Cl→Ar			Cl→Pr			Cl→Rw			Avg (first 6 tasks)		
		OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS
Cluster	✗	6.0	63.0	11.0	5.0	55.0	10.0	4.0	62.0	8.0	8.0	59.0	14.0	5.0	55.0	9.0	5.0	56.0	8.0	5.5	58.3	10.0
SHOT	✓	67.0	28.0	39.5	81.8	26.3	39.8	87.5	32.1	47.0	66.8	46.2	54.6	77.5	27.2	40.2	80.0	25.9	39.1	**76.8**	31.0	43.4
AaD	✓	50.5	67.4	57.7	64.0	66.4	65.1	72.2	69.5	70.8	47.1	80.3	59.3	64.7	68.2	66.4	65.0	71.0	67.8	60.6	70.5	64.5
Ours(w/o s)	✓	47.6	69.2	56.4	70.0	81.8	75.4	70.6	85.6	77.4	60.3	74.7	66.7	65.6	82.7	73.2	72.1	84.2	77.6	64.4	**79.7**	<u>71.1</u>
Ours (w/ s)	✓	49.0	64.8	55.8	68.1	88.0	76.7	71.5	86.6	78.4	61.3	72.6	66.4	66.5	81.6	73.1	71.6	84.8	77.6	<u>64.7</u>	**79.7**	**71.3**
method	SF	Pr→Ar			Pr→Cl			Pr→Rw			Rw→Ar			Rw→Cl			Rw→Pr			Avg (all tasks)		
		OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS	OS*	UNK	HOS
Cluster	✗	8.0	53.0	15.0	6.0	50.0	10.0	5.0	61.0	9.0	8.0	51.0	13.0	5.0	64.0	10.0	5.0	50.0	10.0	5.8	56.6	10.6
SHOT	✓	66.3	51.1	57.7	59.3	31.0	40.8	85.8	31.6	46.2	73.5	50.6	59.9	65.3	28.9	40.1	84.4	28.2	42.3	**74.6**	33.9	45.6
AaD	✓	46.9	83.1	60.0	45.0	72.6	55.6	69.0	72.3	70.6	56.0	77.4	65.0	48.3	67.6	56.4	67.7	69.3	68.5	58.0	72.1	63.6
Ours(w/o s)	✓	60.0	76.7	67.3	52.6	66.2	58.6	70.7	85.1	77.2	56.5	83.3	67.4	49.1	70.7	57.9	66.1	82.1	73.2	<u>61.8</u>	<u>78.5</u>	<u>69.0</u>
Ours(w/ s)	✓	55.9	85.6	67.6	45.4	70.2	55.1	73.9	83.9	78.6	56.7	84.1	67.8	49.3	74.6	59.4	69.5	80.0	74.4	61.6	**79.7**	**69.2**

Table 2. Accuracy (%) on Office-Home for OS-SFDA. Note that 'Ar→Cl' represents that the model per-trained on the source domain Art transfers to the target domain Clipart. The representations of OS*, UNK, HOS, 'w/o s', and 'w/ s' are the same with Tab. 1.

trast to methods such as AaD and SHOT, where only one indicator exhibits competitive performance. SHOT tends to excel in OS*, while AaD typically performs well on UNK. In contrast, our method displays a well-balanced performance across different measurement dimensions, achieving competitive results overall. For the more challenging Office-Home, our method displays more apparent superiority, which achieves 61.6%, 79.7%, and 69.2% average accuracy for OS*, UNK, and HOS, with 3.6%, 7.6%, and 5.6% relative improvement compared with AaD. Furthermore, our method outperforms all no-parameter clustering methods (FINCH, DenMune, and COMIC) for ACC and K on all tasks in Tab. 3 and Tab. 4. The above results demonstrate the superiority and effectiveness of our method.

Impact of Exploration for the Target Class Space. Compared with OS-SFDA methods (SHOT and AaD) that group all target unknown classes as one, our method tends to explore class space and excavate different unknown classes, which has great application value in more realistic scenarios. The experimental results show that our method achieves excellent improvement over the state-of-the-art method AaD for the known accuracy OS* on the more challenging Office Home. In addition, the unknown accuracy UNK for our method is much higher than SHOT on Office-31 and AaD on Office-Home. This phenomenon shows that our method can improve knowledge transfer on known classes when exploring class space and simultaneously improve generalization on unknown classes. To more fully demonstrate the effectiveness and advancement of our exploration, we compare our method with classical no-parameter clustering methods for exploring the class space. Concretely, ours provides the more precise inferred target class number K and higher accuracy ACC on all tasks, where the precision of K represents the efficacy in exploring the target class space, as discussed in Sec. 3.1. The above analysis conclusively demonstrates that our method of exploring class space is exceptionally effective, improving knowledge transfer and generalization.

Effect of Reliable Known Knowledge as Supervision. As introduced in Sec. 3.2, we hope to utilize the reliable known knowledge to optimize the unknown diffuser for better knowledge transfer and generalization. Concretely, our method that uses the reliable known knowledge(ours (w/ s)) outperforms all compared methods(including ours(w/o s)) on more challenging Office-Home for average HOS, ACC, and K in Tabs. 1 to 4. Therefore, the more effectiveness brought by reliable known knowledge ('ours (w/ s)') confirms it guides the model for better known and unknown



(a) (b) (c) (d)

Figure 3. (a) t-SNE of SHOT on R→P. (b) t-SNE of AaD on R→P. (c) t-SNE of our method on R→P. (d) Parameter analysis on A→W.

Methods	A→D	A→W	D→A	D→W	W→A	W→D
FINCH	31.0 (8)	8.0 (2)	23.0 (7)	12.0 (3)	28.0 (9)	23.0 (5)
DenMune	20.0(5)	52.0(15)	58.0(29)	46.0(14)	65.0(35)	22.0(5)
COMIC	56.8 (22)	47.3 (24)	16.1 (40)	46.4 (20)	12.1 (32)	56.6 (22)
Ours(w/o s)	69.0(24)	80.6(28)	60.4(37)	82.5(30)	59.6(38)	68.9(27)
Ours(w/ s)	71.2(25)	73.3(28)	63.1(37)	76.3(32)	64.2(38)	74.5(28)

Table 3. Accuracy ACC(%) and the inferred target class number K on Office-31. Note that the values before '( )' and inside '( )' represent the performance ACC and K, respectively.

Methods	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw
FINCH	5.0(4)	7.0(3)	6.0(3)	8.0(3)	7.0(4)	7.0(4)
DenMune	15.0(12)	58.0(45)	54.0(39)	16.0(7)	55.0(40)	40.0(32)
COMIC	6.9(65)	50.3(92)	9.43(34)	4.0(1)	10.4(51)	2.3(2)
Ours(w/o s)	35.5(69)	60.5(77)	59.1(77)	47.3(61)	58.9(74)	55.6(60)
Ours(w/ s)	35.9(69)	61.4(80)	61.3(77)	46.9(62)	58.9(78)	61.0(60)
Methods	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Pr	Rw→Cl
FINCH	12.0(4)	5.0(4)	4.0(2)	11.0(4)	12.0(7)	7.0(4)
DenMune	16.0(7)	7.0(6)	52.0(37)	26.0(13)	9.0(8)	62.0(43)
COMIC	4.2(2)	3.7(24)	5.7(14)	4.1(1)	5.2(43)	13.3(65)
Ours(w/o s)	48.5(51)	37.2(56)	63.6(75)	38.1(28)	21.6(50)	50.4(75)
Ours(w/ s)	47.8(51)	35.5(56)	64.0(68)	44.2(35)	21.9(50)	51.8(75)

Table 4. ACC(%) and K on Office-Home.

class discrimination, achieving better knowledge transfer on known classes and generalization on unknown classes.

Parameter Analysis. We study the impact of supervision information on unknown diffuser optimization during training by applying a variable parameter to the supervision loss. As shown in Fig. 3(d), the chart indicates that when there is no supervision information, the optimization cannot be completed well to obtain a favorable transfer effect on known or unknown classes. As the parameters increase within a certain range, UNK and HOS also improve; but parameters that are too large will lead to poor optimization results. In addition, it can be seen that OS* is not sensitive to this parameter, and its change trend is relatively stable. Therefore, we should select appropriate high-confidence known pseudo-labels as supervision information to prevent model overfitting for better guiding the optimization of unknown diffuser.

Effectiveness of the Unknown Diffuser. Tabs. 1 to 4 show our superior performance in exploring unknown class space and domain adaptation in OS-SFDA. The superiority in unknown class exploration indicates that our unknown diffuser can excavate potential unknown classes under the guidance of current discriminative class distributions based on the Metropolis-Hastings (M-H) framework. The impressive domain adaptation effects based on the explored wider space also demonstrate the necessity of using the unknown diffuser for unknown class exploration.

Visualization of features. As shown in Fig. 3(a)-(c), t-SNE visualizations indicate that compared with existing methods, our method explores a more accurate and extensive class space, exhibiting remarkable class discrimination.

5. Conclusion

In this paper, we propose a domain adaptation network in OS-SFDA for exploring unknown classes, thereby jointly benefiting known knowledge transfer and unknown generalize. We introduce an unknown diffuser into the source pre-trained model to excavate a wider target class space. And, the reliable known class knowledge and clustered pseudo-labels over the wider class space are captured to serve as supervision during model optimization, realizing impressive knowledge transfer and generalization. Extensive experiments verify the superiority of our method.

6. Acknowledgements

This work is supported in part by the National Key Research and Development Program of China (No. 2023YFC3305600), Joint Fund of Ministry of Education of China (8091B022149, 8091B02072404), National Natural Science Foundation of China (62132016, 62171343, and 62201436), and Fundamental Research Funds for the Central Universities (ZDRC2102).

References

[1] Mohamed Abbas, Adel El-Zoghabi, and Amin Shoukry. Denmune: Density peak based clustering using mutual nearest neighbors. *Pattern Recognit.*, 109:107589, 2021. 3, 6

[2] Silvia Bucci, Mohammad Reza Loghmani, and Tatiana Tommasi. On the effectiveness of image rotation for open set domain adaptation. In *Proc. Eur. Conf. Comput. Vis.*, pages 422–438. Springer, 2020. 2

[3] Jason Chang and John W Fisher III. Parallel sampling of dp mixture models using sub-cluster splits. *Proc. Adv. Neural Inf. Process. Syst.*, 26, 2013. 4

[4] Ning Ding, Yixing Xu, Yehui Tang, Chao Xu, Yunhe Wang, and Dacheng Tao. Source-free domain adaptation via distribution estimation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 7212–7222, 2022. 2

[5] Yuqi Fang, Pew-Thian Yap, Weili Lin, Hongtu Zhu, and Mingxia Liu. Source-free unsupervised domain adaptation: A survey. *arXiv preprint arXiv:2301.00265*, 2022. 1

[6] Zeyu Feng, Chang Xu, and Dacheng Tao. Open-set hypothesis transfer with semantic consistency. *IEEE Trans. Image Process.*, 30:6473–6484, 2021. 1, 2

[7] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.*, 17(1):2096–2030, 2016. 1

[8] W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. 1970. 4, 5

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 770–778, 2016. 6

[10] JoonHo Jang, Byeonghu Na, Dong Hyeok Shin, Mingi Ji, Kyungwoo Song, and Il-Chul Moon. Unknown-aware domain adversarial learning for open-set domain adaptation. *Proc. Adv. Neural Inf. Process. Syst.*, 35:16755–16767, 2022. 2

[11] Taotao Jing, Hongfu Liu, and Zhengming Ding. Towards novel target discovery through open-set domain adaptation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 9322–9331, 2021. 2, 3

[12] Harold W Kuhn. The hungarian method for the assignment problem. *Nav. Res. Logist. Q.*, 2(1-2):83–97, 1955. 6

[13] Jogendra Nath Kundu, Naveen Venkat, Ambareesh Revanur, R Venkatesh Babu, et al. Towards inheritable models for open-set domain adaptation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 12376–12385, 2020. 2

[14] Jonghyun Lee, Dahuin Jung, Junho Yim, and Sungroh Yoon. Confidence score for source-free unsupervised domain adaptation. In *Proc. Int. Conf. Mach. Learn.*, pages 12365–12377. PMLR, 2022. 2, 3

[15] Wuyang Li, Jie Liu, Bo Han, and Yixuan Yuan. Adjustment and alignment for unbiased open set domain adaptation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 24110–24119, 2023. 2

[16] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *Proc. Int. Conf. Mach. Learn.*, pages 6028–6039. PMLR, 2020. 1, 2, 3, 6

[17] Jian Liang, Dapeng Hu, Jiashi Feng, and Ran He. Umad: Universal model adaptation under domain and category shift. *arXiv preprint arXiv:2112.08553*, 2021. 2

[18] Hong Liu, Zhangjie Cao, Mingsheng Long, Jianmin Wang, and Qiang Yang. Separate to adapt: Open set domain adaptation via progressive separation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 2927–2936, 2019. 2, 6

[19] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Unsupervised domain adaptation with residual transfer networks. *Proc. Adv. Neural Inf. Process. Syst.*, 29, 2016. 1

[20] Mingsheng Long, Yue Cao, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Transferable representation learning with deep adaptation networks. *IEEE Trans. Pattern Anal. Mach. Intell.*, 41(12):3071–3085, 2018. 1

[21] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. *Proc. Adv. Neural Inf. Process. Syst.*, 31, 2018. 1

[22] Yadan Luo, Zijian Wang, Zhuoxiao Chen, Zi Huang, and Mahsa Baktashmotlagh. Source-free progressive graph learning for open-set domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2023. 1, 2, 3

[23] Yingwei Pan, Ting Yao, Yehao Li, Chong-Wah Ngo, and Tao Mei. Exploring category-agnostic clusters for open-set domain adaptation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 13867–13875, 2020. 3

[24] 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. 6

[25] Xi Peng, Zhenyu Huang, Jiancheng Lv, Hongyuan Zhu, and Joey Tianyi Zhou. Comic: Multi-view clustering without parameter selection. In *Proc. Int. Conf. Mach. Learn.*, pages 5092–5101. PMLR, 2019. 3, 6

[26] Masoud Pourreza. Open-set source-free domain adaptation in fundus images analysis. 2023. 2

[27] Subhankar Roy, Martin Trapp, Andrea Pilzer, Juho Kannala, Nicu Sebe, Elisa Ricci, and Arno Solin. Uncertainty-guided source-free domain adaptation. In *Proc. Eur. Conf. Comput. Vis.*, pages 537–555. Springer, 2022. 2

[28] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *Proc. Eur. Conf. Comput. Vis.*, pages 213–226. Springer, 2010. 6

[29] Kuniaki Saito and Kate Saenko. Ovanet: One-vs-all network for universal domain adaptation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 9000–9009, 2021. 2

[30] Kuniaki Saito, Shohei Yamamoto, Yoshitaka Ushiku, and Tatsuya Harada. Open set domain adaptation by backpropagation. In *Proc. Eur. Conf. Comput. Vis.*, pages 153–168, 2018. 2

[31] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, and Kate Saenko. Universal domain adaptation through self supervision. *Proc. Adv. Neural Inf. Process. Syst.*, 33:16282–16292, 2020. 1

[32] Saquib Sarfraz, Vivek Sharma, and Rainer Stiefelhagen. Efficient parameter-free clustering using first neighbor relations. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 8934–8943, 2019. 3, 6

[33] Tasfia Shermin, Guojun Lu, Shyh Wei Teng, Manzur Murshed, and Ferdous Sohel. Adversarial network with multiple classifiers for open set domain adaptation. *IEEE Trans. Multimed.*, 23:2732–2744, 2020. 2

[34] Baochen Sun, Jiashi Feng, and Kate Saenko. Return of frustratingly easy domain adaptation. In *Proc. AAAI Conf. Artif. Intell.*, 2016. 1

[35] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 7167–7176, 2017. 1

[36] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 5018–5027, 2017. 6

[37] Guillaume Vray, Devavrat Tomar, Behzad Bozorgtabar, and Jean-Philippe Thiran. Source-free open-set domain adaptation for histopathological images via distilling self-supervised vision transformer. *arXiv preprint arXiv:2307.04596*, 2023. 1, 2, 3

[38] Qian Wang, Fanlin Meng, and Toby P Breckon. Progressively select and reject pseudo-labelled samples for open-set domain adaptation. *arXiv preprint arXiv:2110.12635*, 2021. 2

[39] Joe H Ward Jr. Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.*, 58(301):236–244, 1963. 4, 5

[40] Kun Wei, Xu Yang, Zhe Xu, and Cheng Deng. Class-incremental unsupervised domain adaptation via pseudo-label distillation. *IEEE Trans. Image Process.*, 2024. 1

[41] Haifeng Xia, Handong Zhao, and Zhengming Ding. Adaptive adversarial network for source-free domain adaptation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 9010–9019, 2021. 2

[42] Jiexi Yan, Zhihui Yin, Erkun Yang, Yanhua Yang, and Heng Huang. Learning with diversity: Self-expanded equalization for better generalized deep metric learning. In *Proc. IEEE Int. Conf. Comput. Vis.*, pages 19365–19374, 2023. 2

[43] Shiqi Yang, Yaxing Wang, Joost Van De Weijer, Luis Herranz, and Shangling Jui. Generalized source-free domain adaptation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 8978–8987, 2021. 2

[44] Shiqi Yang, Shangling Jui, Joost van de Weijer, et al. Attracting and dispersing: A simple approach for source-free domain adaptation. *Proc. Adv. Neural Inf. Process. Syst.*, 35: 5802–5815, 2022. 2, 3, 6

[45] Shiqi Yang, Yaxing Wang, Luis Herranz, Shangling Jui, and Joost van de Weijer. Casting a bait for offline and online source-free domain adaptation. *Comput. Vis. Image Underst.*, page 103747, 2023. 2

[46] Xu Yang, Cheng Deng, Tongliang Liu, and Dacheng Tao. Heterogeneous graph attention network for unsupervised multiple-target domain adaptation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(4):1992–2003, 2020. 1

[47] Yixin Zhang, Zilei Wang, and Weinan He. Class relationship embedded learning for source-free unsupervised domain adaptation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pages 7619–7629, 2023. 2, 3

[48] Han Zhao, Xu Yang, and Cheng Deng. Parameter-agnostic deep graph clustering. *ACM Trans. Knowl. Discov. Data*, 18 (3):1–20, 2024. 3

[49] Jingyu Zhuang, Ziliang Chen, Pengxu Wei, Guanbin Li, and Liang Lin. Open set domain adaptation by novel class discovery. *arXiv preprint arXiv:2203.03329*, 2022. 3