A Separation and Alignment Framework for Black-Box Domain Adaptation

Mingxuan Xia[1,2], Junbo Zhao[2], Gengyu Lyu[3], Zenan Huang[2], Tianlei Hu[2], Gang Chen[2], Haobo Wang[1,2*]

[1]School of Software Technology, Zhejiang University
[2]State Key Laboratory of Blockchain and Data Security, Zhejiang University
[3]Faculty of Information Technology, Beijing University of Technology
{xiamingxuan, j.zhao, lccurious, htl, cg, wanghaobo}@zju.edu.cn,
lyugengyu@bjut.edu.cn

Abstract

Black-box domain adaptation (BDA) targets to learn a classifier on an unsupervised target domain while assuming only access to black-box predictors trained from unseen source data. Although a few BDA approaches have demonstrated promise by manipulating the transferred labels, they largely overlook the rich underlying structure in the target domain. To address this problem, we introduce a novel separation and alignment framework for BDA. Firstly, we locate those well-adapted samples via loss ranking and a flexible confidence-thresholding procedure. Then, we introduce a novel graph contrastive learning objective that aligns under-adapted samples to their local neighbors and well-adapted samples. Lastly, the adaptation is finally achieved by a nearest-centroid-augmented objective that exploits the clustering effect in the feature space. Extensive experiments demonstrate that our proposed method outperforms best baselines on benchmark datasets, e.g. improving the averaged per-class accuracy by 4.1% on the VisDA dataset. The source code is available at: https://github.com/MingxuanXia/SEAL.

Introduction

Unsupervised Domain adaptation (UDA) (Ganin and Lempitsky 2015; Saito et al. 2018; Wilson and Cook 2020) is a popular research topic in deep learning that enables models trained on source domains to generalize well to a target domain without accessing target label information. To handle UDA, researchers have developed a plethora of practical algorithms, including adversarial learning-based approaches (Tzeng et al. 2015; Ganin and Lempitsky 2015; Ganin et al. 2016; Tzeng et al. 2017), discrepancy-based methods (Tzeng et al. 2014; Sun and Saenko 2016; Long et al. 2017; Zhuang et al. 2018) and so on. Thanks to these methodological innovations, UDA has been applied to many real-world scenarios such as image classification, speech recognition, and medical data analysis (Long et al. 2016; Kang et al. 2019; Sun et al. 2017; Perone et al. 2019) where obtaining labeled data in a target domain can be costly or impractical.

Though widely applied in many fields, existing UDA methods draw some crucial concerns. Specifically, UDA requires accessing to source data which potentially threatens data privacy and security. Besides, when the source domain contains a large amount of data, practicing these UDA methods may suffer from huge storage and computational costs. Therefore, a new kind of method called Source Free Domain Adaptation (SFDA) (Liang, Hu, and Feng 2020; Li et al. 2020; Yang et al. 2021) is proposed, where source data is not accessed along the adaptation process. Even so, the problem of source information privacy and the limitation of computational resources still remains, due to these methods highly rely on the white-box source model. To mitigate this problem, a special case of SFDA called black-box domain adaptation (BDA) has been proposed recently (Zhang et al. 2021; Liang et al. 2022a), where only a source API can be utilized to train the unlabeled target domain. These methods fully detach the target domain from the source data and source model, which completely mitigates the issues of privacy and computational resources.

To date, few efforts have been made to address the BDA problem. In particular, these algorithms (Zhang et al. 2021; Liang et al. 2022a; Yang et al. 2023) mostly assume many target samples share features with the source domain and thus, are easier to adapt. Inspired by research on learning with noisy labels (LNL), they identify these easy samples to guide target model training. For example, DINE (Liang et al. 2022a) develops adaptive label smoothing for easy and hard samples according to predicted values. (Zhang et al. 2021) selects samples with low loss rank and integrates self-training methods to adapt the remaining data. The state-of-the-art BETA method (Yang et al. 2023) also detects small loss samples which is achieved by Gaussian Mixture Models, and conducts semi-supervised learning after domain division. Besides, mixup techniques (Zhang et al. 2018) are integrated into DINE and BETA that exploit structural input information for boosted performance. Despite the promise, they suffer two main limitations. First, these methods mostly rely on a fixed budget of easy samples but overlook a rich set of well-adapted samples that emerge when the domain gap is shrunk. Second, a simple mixup technique is not able to fully utilize the topological information in the target domain.

To address these problems, we present a novel **SE**paration and **AL**ignment (dubbed **SEAL**) framework for BDA, which comprises three key components. The first module separates target data into *well-adapted* and *under-adapted* sets. Specifically, apart from the original easy samples with lower loss rank, we further collect those well-adapted samples with high confidence by a flexible thresholding mechanism. The

*Corresponding author.

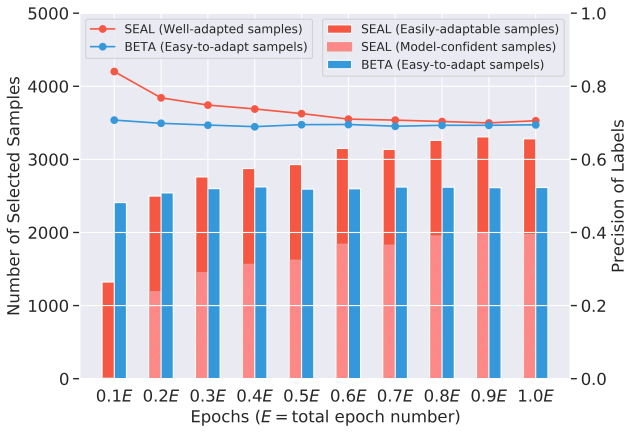

Figure 1: Comparison of the selected number (bar) and the label precision (line) of Well-adapted/Easy-to-adapt samples between SEAL and BETA on Office-Home A→C. SEAL retrieves far more samples while maintaining high precision.

second one is a graph contrastive learning (GCL) module that enforces those under-adapted samples to align their local neighbors as well as the well-adapted peers. As a result, samples in a local region can share similar labels to form distinguishable data clusters. Finally, we devise a nearest-centroid-augmented objective to further exploit the clustering effect in the feature space.

In summary, SEAL utilizes rich topological information to encourage the under-adapted samples to be aligned to the easy domain. With tighter clustering, our separation module is able to excessively collect more well-adapted samples for better training. As shown in Figure 1, SEAL progressively identifies more well-adapted samples with higher label precision, thereby facilitating the adaptation process and attaining superior performance. We evaluate SEAL on various domain adaptation datasets, where SEAL achieves state-of-the-art. Notably, for the VisDA dataset, SEAL outperforms the best baseline by **4.1**% on averaged class-wise accuracy, which validates the effectiveness of the proposed method.

Proposed Method

In this section, we elaborate our novel **SE**paration and **AL**ignment (**SEAL**) framework for black-box domain adaptation (BDA), which excessively collects the well-adapted samples in the target domain and aligns the under-adapted samples by taking in-depth consideration for the underlying feature structure in the target domain. Specifically, we first introduce the domain separation module which detects the easily-adaptable and model-confident peers via small-loss and high-confidence criteria. Then, a novel target domain alignment mechanism is proposed to align the under-adapted samples to their local neighborhood samples and the well-adapted samples, which is accomplished by a graph contrastive learning objective. Finally, we further leverage the fully exploited feature space to provide additional adaptation guidance through nearest-centroid classification. With locality alignment, the aligned samples progressively expand the well-adapted subdomain and achieve domain adaptation. Figure 2 demonstrates the overall framework of SEAL.

Notations. We consider the BDA problem with n source domains $\{\mathcal{X}_{S_j} \times \mathcal{Y}\}_{j=1}^{n}$ and a target domain $\{\mathcal{X}_T \times \mathcal{Y}\}$, where $\mathcal{X}_{S_1}, \mathcal{X}_{S_2}, ..., \mathcal{X}_{S_n}, \mathcal{X}_T$ denote input spaces from different domains and $\mathcal{Y} = \{1, 2, ..., L\}$ denotes the same label space across domains. The goal of BDA is to learn a mapping model $h_T : \mathcal{X}_T \to \mathbb{R}^L$ in the target domain. Notably, the model can only access n black-box predictors $\{h_{S_j}\}_{j=1}^{n}$ and the unlabeled target dataset $\mathcal{D}_T = \{\boldsymbol{x}_T^i\}_{i=1}^{N_T}$ with N_T unlabeled samples during training. $h_{S_j} : \mathcal{X}_{S_j} \to \mathbb{R}^L$ is the source model trained by j^{th} source dataset $\mathcal{D}_{S_j} = \{(\boldsymbol{x}_{S_j}^i, y_{S_j}^i)\}_{i=1}^{N_j}$ with N_j labeled samples and $\boldsymbol{x}_{S_j}^i \in \mathcal{X}_{S_j}, y_{S_j}^i \in \mathcal{Y}, \boldsymbol{x}_T^i \in \mathcal{X}_T$. Hence, the only information we can use from the source domains is the soft predictions of black-box source models on the target domain, $\boldsymbol{z}_{S_j}^i = h_{S_j}(\boldsymbol{x}_T^i)$. By fairly treating all source domains, we can obtain a unified source-predicted soft label $\boldsymbol{z}_i$ for each target sample $\boldsymbol{x}_i$ by computing the mean value of all predictions as $\boldsymbol{z}_i = \frac{1}{n}\sum_{j=1}^{n} h_{S_j}(\boldsymbol{x}_i)$.

Well/Under-Adapted Domain Separation

To begin with, we describe the domain separation step. Following previous works (Liang et al. 2022a; Yang et al. 2023), we assume there exists a shared subdomain between the source and target domain. Subsequently, most existing algorithms first fit such a simple subset and then handle the remaining data adaptation. In our work, we observe that this decent set is not set in stone. As the domain gap narrows, more data can be aligned to the easily adaptable region. The typical characteristic of these new samples is that the model's predictions become quite confident. Motivated by this, we separate the whole target domain into two subdomains, namely the **well-adapted subdomain** $\mathcal{D}^W$ (**WD**) and the **under-adapted subdomain** $\mathcal{D}^U$ (**UD**). In particular, we expect WD to contain two kinds of samples, namely the original *easily-adaptable* samples and *model-confident* samples.

Easily-adaptable Sample Selection. The deep networks are known to pose the memorization effect (Zhang et al. 2017) and are prone to fit easy patterns prior. Therefore, easily adaptable samples can dominate the training process and produce lower loss ranks at the early stage of training (Zhang et al. 2021). Motivated by this, we compute the cross-entropy loss for each sample by its source-predicted label (Onehot form) and the model's outputs and then select the top-δ percent samples with the smallest loss. Formally, we denote the easily-adaptable part of WD as:

$$\mathcal{D}^{W_e} = \{\boldsymbol{x}_i | l_i \in \mathcal{C}^\delta, l_i = \mathcal{H}(\hat{\boldsymbol{z}}_i, \boldsymbol{p}_i)\} \tag{1}$$

where $\mathcal{H}(\cdot, \cdot)$ denotes the cross-entropy loss, $\hat{\boldsymbol{z}}_i$ denotes the Onehot form of $\boldsymbol{z}_i$, $\boldsymbol{p}_i = h_T(\boldsymbol{x}_i)$ denotes output logits of the target model, and $\mathcal{C}^\delta$ denotes the top-δ percent smallest losses. In SEAL, this subset serves as a pivot that helps the model initialize and guides domain alignment.

Model-confident Sample Selection. When trained on easily-adaptable samples, the classifier naturally clusters them tightly to their respective class centers. Alongside model

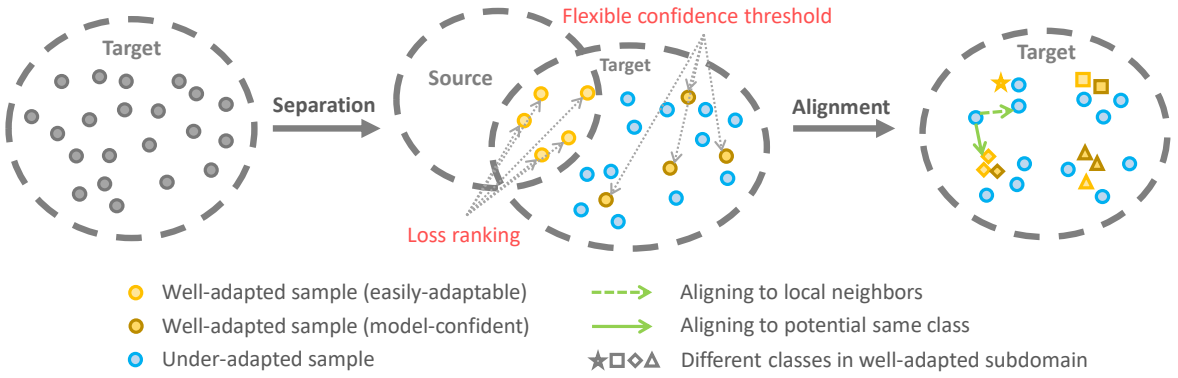

Figure 2: The overall framework of SEAL. The target domain is first separated into well/under-adapted subdomain, where the well-adapted subdomain consists of easily-adaptable samples with small loss rank and model-confident samples with high confidence prediction. Then, domain alignment is conducted to pull under-adapted peers close to those peers in the local neighborhood and with the potential same class. Furthermore, the alignment enhances the cluster effect, which encourages the under-adapted samples to perform confident prediction and consequently enlarges the well-adapted subdomain.

training, SEAL attempts to align all samples to a coherent domain. As a result, more samples can be absorbed into the clusters. Notably, it is not suitable to reuse the loss-based selection protocol since these newly adapted samples can still pose a high loss rank. However, these samples typically produce higher confidence. Motivated by this, SEAL selects these newly adapted samples via a flexible confidence thresholding mechanism that considers the adaptation progress for different classes in different stages. Formally, for class c at epoch e, we calculate the following per-class threshold for sample selection,

$$\mathcal{T}_e(c) = \hat{a}_e(c) \cdot \tau \tag{2}$$

where τ is the pre-defined high confidence threshold. $\hat{a}_e(c)$ represent the adaptation progress of class c at epoch t, which indicates how well a class is adapted. We estimate the adaptation progress by the number of samples in the whole target domain that perform a high-confident output, i.e., whose maximum logits are above the pre-defined high threshold. We then formalize the adaptation progress as follows:

$$a_e(c) = \sum_{\boldsymbol{x}_i \in \mathcal{D}_T} \mathbb{I}(\max(\boldsymbol{p}_i) > \tau) \cdot \mathbb{I}(\arg\max(\boldsymbol{p}_i) = c),$$
$$\text{where} \quad \hat{a}_e(c) = \frac{a_e(c)^\gamma}{\max a_e^\gamma} \tag{3}$$

where the adaptation progress a_e is scaled and normalized to $\hat{a}_e$, and γ is a hyper-parameter ranging from 0 to 1. In other words, we expect those easier classes will have more samples above τ. Then, we set normalized and scaled counts to flexibly adjust the threshold for different classes.

Remark. It should be noted that the adaptation procedure may face inherent imbalance in the target domain due to label skewness or inherent difficulty in adapting some classes. Therefore, a fixed threshold can be problematic. With a high threshold, those hard classes may never be selected; with a too-low threshold, we may excessively select under-adapted samples. Fortunately, even though the model prediction across different classes is not comparable, newly-adapted samples can still be closer to the class centers than under-adapted ones in the same class. To achieve this, our flexible thresholding mechanism leverages the whole dataset statistics to estimate the adaptation progress and additionally scales it to confront the imbalance problem.

Finally, the model-confident part of WD can be formalized as:

$$\mathcal{D}^{W_c} = \{\boldsymbol{x}_i | \max(\boldsymbol{p}_i) > \mathcal{T}_e(\arg\max(\boldsymbol{p}_i)\} \tag{4}$$

Overall, the well-adapted subdomain is organized by combining the easily-adaptable part and the model-confident part as $\mathcal{D}^W = \mathcal{D}^{W_e} \cup \mathcal{D}^{W_c}$, and the rest samples compose the under-adapted subdomain, i.e., $\mathcal{D}^U = \mathcal{D} - \mathcal{D}^W$.

Graph Contrastive Learning for Locality Alignment

In this section, we introduce how we promote adaptation for the under-adapted subdomain. Different from previous work (Liang et al. 2022a; Yang et al. 2023) which guides the adaptation in a label-driven manner, we aim to align the under-adapted samples via feature structure where data intrinsically forms clusters, which further avoids accessing unreliable label information of UD. Specifically, we align under-adapted samples not only to well-adapted samples with the same potential classes but also to their local neighbors and their own augmented view in feature space. To achieve the above alignments, we devise a graph contrastive learning objective via a locality label graph and a feature graph.

To begin with, we first introduce the model architecture of our framework. For each target sample $\boldsymbol{x} \in \mathcal{D}_T$, we first generate a weakly-augmented view $\alpha(\boldsymbol{x})$ which is forwarded to the target model h_T to obtain classification results. Additionally, a projection network $g(\cdot)$ sharing the same encoder with classifier h_T is jointly trained, which embeds each weakly-augmented sample into a normalized low-dimensional feature vector $\boldsymbol{q} \in \mathbb{R}^d$. We further generate a strongly-augmented

view $\mathcal{A}(\boldsymbol{x})$ and also embed it into d-dimensional feature vector $\boldsymbol{q}' = g'(\mathcal{A}(\boldsymbol{x}))$, where $g'(\cdot)$ is the momentum updated version of the projection network $g(\cdot)$. In implementation, we use SimAugment (Khosla et al. 2020) as the weak augmentation module and RandAugment (Cubuk et al. 2019) as the strong augmentation module. Note that our domain separation mechanism works on original images without gradient. Then, given a batch of samples $\{\boldsymbol{x}_i\}_{i=1}^B$, we construct the locality label graph $\boldsymbol{W}^L$ of size $B \times B$ as:

$$W_{ij}^L = \begin{cases} \mathbb{I}(\hat{y}_i = \hat{y}_j), & \boldsymbol{x}_i \in \mathcal{D}^W \wedge \boldsymbol{x}_j \in \mathcal{D}^W \\ \lambda \mathbb{I}(\hat{y}_i = \hat{y}_j), & \boldsymbol{x}_i \in \mathcal{D}^W \wedge \boldsymbol{x}_j \in \mathcal{D}^U \\ \lambda \mathbb{I}(\hat{y}_i = \hat{y}_j \vee \boldsymbol{q}_j \in \mathcal{N}_i^k), & \boldsymbol{x}_i \in \mathcal{D}^U \end{cases} \tag{5}$$

where $\mathcal{N}_i^k$ denotes the feature set of q_i's k-nearest neighbors. $\hat{y}$ denotes the potential class of each sample, where we take the source-predicted label for samples in $\mathcal{D}^{W_e}$ and self-predicted label for other samples, i.e., $\hat{y} = \arg\max \boldsymbol{z}$ if $x \in \mathcal{D}^{W_e}$ else $\hat{y} = \arg\max h_T(\alpha(\boldsymbol{x}))$. λ denotes the weight of the edge in the label graph which is a hyper-parameter ranging between 0 and 1. A lower weight for those edges accessing under-adapted samples enables a more error-tolerant alignment process. We further fill in 1 on the diagonal of $\boldsymbol{W}^L$ which encourages strong self-alignment for all samples. In this way, the locality label graph connects samples in local neighborhoods and potential same classes, which later served as the training target for the feature graph to conduct domain alignment through feature similarity. The feature graph is then constructed via the similarities of features generated by two views, formally:

$$W_{ij}^F = \exp(\boldsymbol{q}_i^\top \cdot \boldsymbol{q}_j'/t) \tag{6}$$

where t is the temperature. Finally, we formalize the objective of our graph contrastive learning as the cross-entropy loss between the feature graph and the label graph:

$$\mathcal{L}_{cont} = \frac{1}{B} \sum_{i=1}^{B} \mathcal{H}(\hat{\boldsymbol{W}}_i^L, \hat{\boldsymbol{W}}_i^F) \tag{7}$$

where $\hat{W}_{ij} = W_{ij} / \sum_j W_{ij}$ is the row-wise normalized version of each graph.

Notably, our graph contrastive objective enjoys two main advantages. First, contrastive learning has demonstrated a great ability in improving representation (He et al. 2020; Khosla et al. 2020), which can largely benefit domain adaption. Second, the basic idea of contrastive learning is to utilize the smoothed augmented-neighborhood property of every single sample for representation learning, while SEAL goes beyond by aligning samples in a larger local neighborhood, which further encourages the exploration of relationships among data clusters.

Overall Framework

As we mentioned above, our graph contrastive learning effectively mines the feature structure and forms tight clusters. We would like to further incorporate feature information into our training objective. Specifically, we conduct a nearest-centroid classification and obtain a soft logits vector $\boldsymbol{r}$ by the scaled softmax of cosine similarity between sample feature $\boldsymbol{q}$ and the class mean feature $\boldsymbol{\mu}_c$:

$$r_c = \frac{\exp(\boldsymbol{q}^\top \boldsymbol{\mu}_c / t)}{\sum_{l=1}^{L} \exp(\boldsymbol{q}^\top \boldsymbol{\mu}_l / t)}, \forall 1 \le c \le L \tag{8}$$

where $\boldsymbol{\mu}_c$ is the average of per-class features according to the model predictions. We then introduce the training target for each sample according to our separation results. For those well-adapted samples, we adopt source-predicted label $\hat{\boldsymbol{z}} = \text{Onehot}(\boldsymbol{z})$ for the easily-adaptable part and model-predicted label $\hat{\boldsymbol{p}} = \text{Onehot}(\boldsymbol{p})$ for the model-confident part. Note that the model's prediction $\boldsymbol{p} = h_T(\boldsymbol{x})$ is obtained by samples without augmentations and with no gradient. We further ensemble nearest-centroid logits $\boldsymbol{r}$ into the training target of the model-confident part and enable this feature information to take more proportion on the refinery of low-confident peers. For under-adapted samples, we directly use the nearest-centroid output as their training target. Formally:

$$\tilde{\boldsymbol{y}} = \begin{cases} \hat{\boldsymbol{z}}, & \text{if } \boldsymbol{x}_i \in \mathcal{D}^{W_e} \\ \max(\boldsymbol{p}) \cdot \hat{\boldsymbol{p}} + (1 - \max(\boldsymbol{p})) \cdot \boldsymbol{r}, & \text{if } \boldsymbol{x}_i \in \mathcal{D}^{W_c} \\ \boldsymbol{r}, & \text{if } \boldsymbol{x}_i \in \mathcal{D}^U \end{cases} \tag{9}$$

Then, we describe the classification loss as the cross-entropy loss between the output of weakly-augmented branches and their corresponding labels. Additionally, to encourage the diversification of model outputs, we further combine an entropy regularizer term which is widely adopted in UDA (Liang, Hu, and Feng 2020; Chen et al. 2022a). Formally, we define the classification loss as:

$$\mathcal{L}_{cls} = \mathcal{H}(\tilde{\boldsymbol{y}}, h_T(\alpha(\boldsymbol{x}))) + \mathcal{R}(\mathbb{E}_{\boldsymbol{x} \in D_T} h_T(\alpha(\boldsymbol{x}))) \tag{10}$$

where $\mathcal{R}(\cdot)$ denotes the entropy operation. Finally, the overall objective is defined by the sum of classification loss and contrastive loss, i.e., $\mathcal{L}_{total} = \mathcal{L}_{cls} + \eta \cdot \mathcal{L}_{cont}$, where η is a hyper-parameter controlling contrastive loss weights. The pseudo-code is summarized in Appendix.

To sum up, the black-box domain adaptation process benefits from our separation and alignment mechanism in the following procedure. Firstly, the well-adapted domain separation module detects easily-adaptable and model-confident samples to enhance classifier training and serve as aligning anchors for under-adapted samples. Then, the graph contrastive learning mechanism induces under-adapted samples aligning to their local neighbors as well as the well-adapted subdomain. This locality alignment also forms tight clusters and further encourages the under-adapted samples to perform confident output through a nearest-centroid objective, which consequently enlarges the well-adapted subdomain.

Experiments

In this section, we report our main empirical results to show the superiority of SEAL. We refer the readers to Appendix for more experimental details and results.

Setup

Datasets. **Office** (Saenko et al. 2010) is the most popular image recognition benchmark for UDA. It contains 31 categories in three domains, namely Amazon, DSLR, and Webcam, with 4,652 images in total. **Office-Home** (Venkateswara

Method	Ar:Cl	Ar:Pr	Ar:Re	Cl:Ar	Cl:Pr	Cl:Re	Pr:Ar	Pr:Cl	Pr:Re	Re:Ar	Re:Cl	Re:Pr	Avg.
Source	44.1	66.9	74.2	54.5	63.3	66.1	52.8	41.2	73.2	66.1	46.7	77.5	60.6
NLL-OT	49.1	71.7	77.3	60.2	68.7	73.1	57.0	46.5	76.8	67.1	52.3	79.5	64.9
NLL-KL	49.0	71.5	77.1	59.0	68.7	72.9	56.4	46.9	76.6	66.2	52.3	79.1	64.6
HD-SHOT	48.6	72.8	77.0	60.7	70.0	73.2	56.6	47.0	76.7	67.5	52.6	80.2	65.3
SD-SHOT	50.1	75.0	78.8	63.2	72.9	76.4	60.0	48.0	79.4	69.2	54.2	81.6	67.4
DINE	52.2	78.4	81.3	65.3	76.6	78.7	62.7	49.6	82.2	69.8	55.8	84.2	69.7
BETA	57.2	78.5	82.1	68.0	78.6	79.7	67.5	**56.0**	83.0	71.9	58.9	84.2	72.1
SEAL	**58.5**	**81.4**	**84.7**	**71.7**	**80.4**	**82.1**	**72.2**	54.3	**86.0**	**76.2**	**60.6**	**86.3**	**74.5**

Table 1: Accuracies (%) on Office-Home for black-box domain adaptation.

Method	A:D	A:W	D:A	D:W	W:A	W:D	Avg.
Source	79.9	76.6	56.4	92.8	60.9	98.5	77.5
NLL-OT	88.8	85.5	64.6	95.1	66.7	98.7	83.2
NLL-KL	89.4	86.8	65.1	94.8	67.1	98.7	83.6
HD-SHOT	86.5	83.1	66.1	95.1	68.9	98.1	83.0
SD-SHOT	89.2	83.7	67.9	95.3	71.1	97.1	84.1
DINE	91.6	86.8	72.2	**96.2**	73.3	98.6	86.4
BETA	93.6	**88.3**	76.1	95.5	76.5	99.0	88.2
SEAL	**95.1**	**88.3**	**77.6**	96.0	**76.7**	**99.3**	**88.8**

Table 2: Accuracies (%) on Office for black-box domain adaptation.

et al. 2017) is a more challenging medium-sized dataset with 65 classes from four domains: Art, Clipart, Product, and Real-World, which contains 15,500 images. **VisDA** (Peng et al. 2017) is a large-scaled synthetic-to-real dataset for computer vision. It contains 12 categories with 152k synthetic images in the source domain and 55k real object images from Microsoft COCO in the target domain. **DomainNet** (Peng et al. 2019) is the largest domain adaptation benchmark with over 596k images from 6 domains in 345 categories.

Baselines. We compare SEAL with the following seven BDA methods: 1) **BETA** (Yang et al. 2023) is the recent state-of-the-art method that divides the target domain into easy and hard subdomains by loss criterion and then conducts semi-supervised learning. 2) **DINE** (Liang et al. 2022a) leverages adaptive self-knowledge distillation alongside structural regularizations and mutual information maximization. 3) **HD-SHOT** is derived from the source-free UDA method SHOT (Liang, Hu, and Feng 2020), which first obtains a model in a self-training manner and then conducts adaptation by SHOT using the source-predicted labels. 4) **SD-SHOT** also applies SHOT and differs from HD-SHOT by using weighted cross-entropy loss. 5) **NLL-KL** is based on noisy label learning (NLL) (Zhang et al. 2021) which leverages KL divergence to purify pseudo labels and trains the model iteratively. 6) **NLL-OT** also follows the idea of NLL and adopts optimal transport (OT) (Asano, Rupprecht, and Vedaldi 2020) to refine the pseudo labels. 7) **Source** is a simple baseline that directly uses the black-box source model for evaluation.

Implementation Details. For a fair comparison, we train black-box source models following the implementation in DINE (Liang et al. 2022a). We use the pre-trained ResNet50 and ResNet101 (He et al. 2016) as the backbone and a linear model with weight normalization as the classifier for target models. A bottleneck with the size of 256 is added between the backbone and classifier, which is widely adopted (Liang, Hu, and Feng 2020; Liang et al. 2022a). The projection head is implemented right after the backbone, which is a 2-layer MLP with an output dimension of 128. The model is trained using a standard SGD optimizer with a learning rate of $\epsilon_0 = 1e^{-3}$ for the backbone and $\epsilon_0 = 1e^{-2}$ for other layers. The momentum, weight decay, and batch size are set to 0.9, $1e^{-3}$, and 64 respectively, and the learning rate schedule is set as $\epsilon = \epsilon_0 \cdot (1 + 10 \cdot \frac{e}{E})^{-0.75}$. The total training epochs E is set to 100 for all datasets except 20 for VisDA and 50 for DomainNet. For hyper-parameters, we fix the high threshold τ, temperature t, and contrastive loss weight η as 0.95, 0.07, and 0.5 respectively. The number of neighbors k is 5/3/100 for Office/Office-Home/VisDA. To enlarge the neighborhood in the locality label graph, we maintain a feature queue that stores the most current M momentum features $\{q'_j\}_{j=1}^{M}$ where M is empirically set as 1024, and we modify the graphs to have size $B \times (B + M)$. Moreover, we implement an extra training stage between every training epoch with self-knowledge distillation and mutual information maximization objective (Liang et al. 2022a). We also add a self-supervised contrastive loss as a warm-up stage to obtain stable features. Furthermore, mixup training (Zhang et al. 2018) and consistency regularization (Sohn et al. 2020) are utilized to improve the robustness of SEAL; see Appendix. For all experiments, we report the mean accuracy from 3 independent runs with different random seeds.

Results

Results on Office and Office-Home. For Office and Office-Home, we report the top-1 accuracy of every domain shift as well as the average top-1 accuracy of all domain shifts (denotes Avg.). The results of Office-Home and Office are shown in Table 1 and Table 2. Overall, SEAL achieves state-of-the-art among all baseline methods, where SEAL improves upon **2.4**% of the average accuracy for Office-Home and **0.6**% for Office. Specifically, SEAL demonstrates the best performance among all domain shifts except for Pr:Cl, and significantly outperforms the best baseline on Pr:Ar by a margin of **4.7**%. For Office, SEAL achieves superior results on most domain shifts, where SEAL improves the best baseline by a margin of **1.5**% on A:D and D:A.

Method	plane	bcycl	bus	car	horse	knife	mcycle	person	plant	sktbrd	train	truck	c-Avg.
Source	64.3	24.6	47.9	75.3	69.6	8.5	79.0	31.6	64.4	31.0	81.4	9.2	48.9
NLL-OT	82.6	84.1	76.2	44.8	90.8	39.1	76.7	72.0	82.6	81.2	82.7	50.6	72.0
NLL-KL	82.7	83.4	76.7	44.9	90.9	38.5	78.4	71.6	82.4	80.3	82.9	50.4	71.9
HD-SHOT	75.8	85.8	78.0	43.1	92.0	41.0	79.9	78.1	84.2	86.4	81.0	65.5	74.2
SD-SHOT	79.1	85.8	77.2	43.4	91.6	41.0	80.0	78.3	84.7	86.8	81.1	65.1	74.5
DINE	81.4	86.7	77.9	55.1	92.2	34.6	80.8	79.9	87.3	87.9	84.3	58.7	75.6
BETA	94.9	90.2	85.4	61.1	95.9	93.1	85.0	83.8	92.9	91.9	91.1	55.0	85.1
SEAL	**97.9**	**92.2**	**88.0**	**73.5**	**97.1**	**96.1**	**92.4**	**85.7**	**93.9**	**95.6**	**91.2**	**66.4**	**89.2**

Table 3: Accuracies (%) on VisDA for black-box domain adaptation.

Method	C:P	C:R	C:S	P:C	P:R	P:S	R:C	R:P	R:S	S:C	S:P	S:R	Avg.
Source	36.1	52.1	41.3	40.7	56.5	34.6	48.3	46.8	35.2	50.5	35.9	46.1	43.7
DINE	43.7	61.5	44.0	44.0	62.9	38.7	54.3	53.1	41.7	54.0	44.5	59.3	50.1
BETA	48.3	64.7	**49.2**	49.6	66.3	43.4	58.1	**57.7**	45.7	58.7	49.9	63.1	54.5
SEAL	**49.5**	**67.9**	48.7	**49.9**	**68.5**	**44.0**	**60.6**	57.4	**46.7**	**59.2**	**50.4**	**67.1**	**55.8**

Table 4: Accuracies (%) on DomainNet for black-box domain adaptation.

Method	Ar	Cl	Pr	Re	Avg.
Source	54.9	49.9	69.6	76.7	62.8
DINE	74.8	**64.1**	85.0	84.6	77.1
SEAL	**77.5**	**64.1**	**86.4**	**87.2**	**78.8**

Table 5: Accuracies (%) on Office-Home for multi-source black-box domain adaptation.

Results on VisDA and DomainNet. For the more challenging large-scale benchmark VisDA, we report the class-wise top-1 accuracy and the averaged class-wise accuracy (denotes c-Avg.). Table 3 demonstrates the comparison results, where SEAL outperforms all the rivals in all classes and significantly improves the best baseline by a margin of **4.1**% of the averaged class-wise accuracy. For the largest dataset DomainNet, we conduct experiments on 12 domain shifts from 4 domains (Clipart, Painting, Real, Sketch), which follows the settings in (Jiang et al. 2022). As shown in Table 4, our method performs the best results with an average improvement of **1.3**% on BETA. These results clearly validate the superiority of our separation and alignment framework.

Results on the multi-source adaptation task. We further conduct experiments of multi-source black-box domain adaptation on Office-Home. We follow the settings in DINE where ResNet101 is used as the backbone for training the source and target domains. Table 5 shows that SEAL improves DINE upon **1.7**% averaged accuracy, indicating the superiority of SEAL in multi-source scenarios. Full results on multi-source benchmarks can be found in the Appendix.

Ablation Studies

In this section, we first present ablation results to validate the effectiveness of different components in our proposed method on VisDA. Then, analyze hyper-parameter sensitivity on and VisDA. More ablation can be found in Appendix.

Module	Method	Acc.	c-Avg.
-	SEAL	86.4	89.2
Separation	SEAL w/o Sepa.	84.3	86.9
	SEAL w/o Conf.	85.2	87.5
Alignment	SEAL w/o $\mathcal{L}_{cont}$	81.8	85.3
	SEAL w/o Neig.	85.2	87.9
	SEAL w/o NCC	84.7	87.1

Table 6: Ablation studies of different components in SEAL on VisDA.

Effect of Separation. We compare SEAL with the following three variants: 1) *SEAL w/o Sepa.* which removes separation stage and only trains classifier by self-distilled objective following warm-up stage; 2) *SEAL w/o Conf.* neglects model-confident sample selection and only uses easily-adaptable samples to form the well-adapted subdomain. As shown in Table 6, SEAL w/o Sepa. suffers from a performance drop of **2.1**% overall accuracy and **2.3**% averaged class-wise accuracy, which verifies the effectiveness of our domain separation strategy. Moreover, SEAL outperforms SEAL w/o Conf. by a margin of **1.7**% on averaged class-wise accuracy, which demonstrates the superiority of model-confident sample selection mechanism where more well-adapted samples are engaged to facilitate the training process.

Effect of Alignment. To verify the effectiveness of locality alignment module, we compare with three variants: 1) *SEAL w/o* $\mathcal{L}_{cont}$ trains SEAL without graph contrastive learning loss; 2) *SEAL w/o Neig.* neglects alignment to neighborhood samples by dropping the neighbor criterion in Eq.(5); 3) *SEAL w/o NCC* removes the module of nearest-centroid classification (NCC) and regard those under-adapted samples as unlabeled samples. From Table 6, firstly, we can observe that SEAL w/o $\mathcal{L}_{cont}$ worsens the performance because it fails to

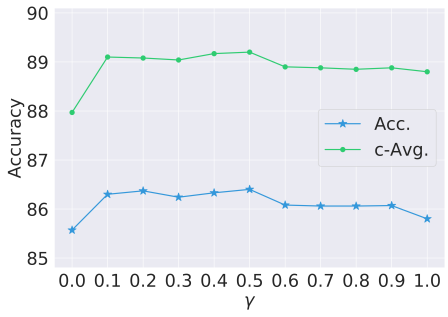

Figure 3: Analysis of parameter γ on VisDA.

align under-adapted samples to well-adapted ones and their neighbors, and consequently fails to form tight clusters. Secondly, SEAL outperforms SEAL w/o Neig. by a margin of **1.3**% averaged accuracy since it ignores the valuable information in local neighborhoods. Finally, SEAL w/o NCC suffers from a performance drop, where under-adapted samples with intrinsic cluster effects are not fully leveraged. The above results indicate that our graph contrastive learning-based locality alignment is indispensable to SEAL when training the under-adapted subdomain.

Parameter Sensitivity. Furthermore, we analyze the effect of scaling parameter γ for model-confident sample selection. As shown in Figure 3, SEAL with $\gamma = 0.5$ achieves the best performance. This observation indicates that, with a proper scaling factor, SEAL can obtain a balanced estimation of the adaptation progress for different classes which either avoids setting the confidence threshold too high or too low, thus excessively selecting model-confident samples that are reliable enough to train a good model. Besides, we observe that SEAL with $\gamma > 0.0$ outperforms the case when $\gamma = 0.0$ where the threshold is fixed as the pre-defined high threshold, which further proves the robustness of γ for our flexible thresholding mechanism. According to the size of different datasets, we empirically set $\gamma = 0.05, 0.1, 0.5, 0.1$ for Office, Office-Home, VisDA, and DomainNet respectively.

Related Work

Domain Adaptation. Unsupervised Domain Adaptation (**UDA**) aims to leverage the knowledge from an abundantly labeled source domain to learn an effective predictor for the target domain with no label information. In recent years, a plethora of deep UDA methods have been proposed to mitigate this domain-shifting problem, including discrepancy minimization methods (Tzeng et al. 2014; Sun and Saenko 2016; Long et al. 2017) and adversarial-based method (Ganin and Lempitsky 2015; Ganin et al. 2016; Tzeng et al. 2017). Other deep UDA approaches tackles the domain discrepancy via pseudo-labeling (Saito, Ushiku, and Harada 2017; Xie et al. 2018; Xiao et al. 2023), domain reconstruction (Vincent et al. 2008; Ghifary et al. 2015), disentanglement representation (Liu et al. 2018), and attention mechanism (Zhuo et al. 2017; Moon and Carbonell 2017). However, these UDA methods highly rely on the source data, which brings social concern about data privacy. To address this problem, Source Free Domain Adaptation (**SFDA**) is proposed where the target domain is adapted without accessing source data. Data generation-based method (Li et al. 2020; Tian et al. 2021; Yang et al. 2022; Ding et al. 2022) synthesizes source data or distribution to apply standard UDA techniques. Model fine-tuning-based methods train the target model initialized by the source model via hidden structure mining (Liang, Hu, and Feng 2020; Yang et al. 2021; Liang et al. 2022b), self-knowledge distillation (Chen et al. 2022b), statistical alignment (Ishii and Sugiyama 2021), or uncertainty guidance (Sivaprasad and Fleuret 2021; Litrico, Del Bue, and Morerio 2023). Nonetheless, the issue of source data privacy and security still remains because the white-box source model is exposed to the target domain. To this end, (Zhang et al. 2021) proposes to select and train the clean target samples with reliable source predictions to handle the Black-box Domain Adaptation (**BDA**) problem. Recent state-of-the-art method (Yang et al. 2023) selects easy-to-adapt samples on a fixed budget and trains the model in a semi-supervised manner with mixup techniques. Different from existing BDA methods, SEAL focuses on separating the well-adapted samples via loss ranking and confidence thresholding, which serve as pivots for the alignment of under-adapted samples.

Contrastive Learning. Contrastive learning (Chen et al. 2020; He et al. 2020; van den Oord, Li, and Vinyals 2018) is proposed to learn discriminative visual representations in a self-supervised manner. Recent work further proposes supervised contrastive learning (Khosla et al. 2020), which pulls clusters of points in the same class together and pushes apart those points from different classes. Motivated by the effectiveness of supervised contrastive learning, researchers adopt contrastive learning to a variety of weakly-supervised tasks, such as semi-supervised learning (Li, Xiong, and Hoi 2021; Ye et al. 2023), partial-label learning (Wang et al. 2022), and noisy-label learning (Wu et al. 2021; Wei et al. 2022), which benefit from the robust cluster information. There are also methods that utilize contrastive learning to mitigate the domain gap for domain adaptation problems (Singh 2021; Thota and Leontidis 2021; Chen et al. 2022a; Wang et al. 2023). Note that these methods work in a white-box setting where the trained source model is leveraged to initialize the target model and the source classifier weights can serve as class prototypes, our proposed method practices contrastive learning without accessing the source model or data. Moreover, apart from aligning features to their self-augmented view, we further align samples in a larger neighborhood.

Conclusion

This work introduces a novel SEAL framework for the black-box domain adaptation problem. SEAL is driven by a separation component for excessively collecting well-adapted samples and a graph contrastive learning module for aligning remaining under-adapted samples. By then, SEAL can uncover underlying structural information to reduce the domain gap. Extensive experiments demonstrate that SEAL improves state-of-the-art BDA algorithms by a large margin. We hope our work will inspire future research to develop powerful BDA algorithms that are aware of topological structures.

Acknowledgments

This work is majorly supported by the National Natural Science Foundation of China under grants (No. 62206247). Junbo Zhao also thanks the sponsorship by the Fundamental Research Funds for the Central Universities (No. 226-2022-00028). Gengyu Lyu would like to thank the National Natural Science Foundation of China under grants (No. 62306020).

References

Asano, Y. M.; Rupprecht, C.; and Vedaldi, A. 2020. Self-labelling via simultaneous clustering and representation learning. In *ICLR*. OpenReview.net.

Chen, D.; Wang, D.; Darrell, T.; and Ebrahimi, S. 2022a. Contrastive Test-Time Adaptation. In *CVPR*, 295–305. IEEE.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. E. 2020. A Simple Framework for Contrastive Learning of Visual Representations. In *ICML*, volume 119 of *Proceedings of Machine Learning Research*, 1597–1607. PMLR.

Chen, W.; Lin, L.; Yang, S.; Xie, D.; Pu, S.; and Zhuang, Y. 2022b. Self-Supervised Noisy Label Learning for Source-Free Unsupervised Domain Adaptation. In *IROS*, 10185–10192. IEEE.

Cubuk, E. D.; Zoph, B.; Shlens, J.; and Le, Q. V. 2019. RandAugment: Practical data augmentation with no separate search. *CoRR*, abs/1909.13719.

Ding, N.; Xu, Y.; Tang, Y.; Xu, C.; Wang, Y.; and Tao, D. 2022. Source-Free Domain Adaptation via Distribution Estimation. In *CVPR*, 7202–7212. IEEE.

Ganin, Y.; and Lempitsky, V. S. 2015. Unsupervised Domain Adaptation by Backpropagation. In *ICML*, volume 37 of *JMLR Workshop and Conference Proceedings*, 1180–1189. JMLR.org.

Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; and Lempitsky, V. S. 2016. Domain-Adversarial Training of Neural Networks. *J. Mach. Learn. Res.*, 17: 59:1–59:35.

Ghifary, M.; Kleijn, W. B.; Zhang, M.; and Balduzzi, D. 2015. Domain Generalization for Object Recognition with Multi-task Autoencoders. In *ICCV*, 2551–2559. IEEE Computer Society.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. B. 2020. Momentum Contrast for Unsupervised Visual Representation Learning. In *CVPR*, 9726–9735. Computer Vision Foundation / IEEE.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *CVPR*, 770–778. IEEE Computer Society.

Ishii, M.; and Sugiyama, M. 2021. Source-free Domain Adaptation via Distributional Alignment by Matching Batch Normalization Statistics. *CoRR*, abs/2101.10842.

Jiang, J.; Shu, Y.; Wang, J.; and Long, M. 2022. Transferability in Deep Learning: A Survey. *CoRR*, abs/2201.05867.

Kang, G.; Jiang, L.; Yang, Y.; and Hauptmann, A. G. 2019. Contrastive Adaptation Network for Unsupervised Domain Adaptation. In *CVPR*, 4893–4902. Computer Vision Foundation / IEEE.

Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised Contrastive Learning. In *NeurIPS*.

Li, J.; Xiong, C.; and Hoi, S. C. H. 2021. CoMatch: Semi-supervised Learning with Contrastive Graph Regularization. In *ICCV*, 9455–9464. IEEE.

Li, R.; Jiao, Q.; Cao, W.; Wong, H.; and Wu, S. 2020. Model Adaptation: Unsupervised Domain Adaptation Without Source Data. In *CVPR*, 9638–9647. Computer Vision Foundation / IEEE.

Liang, J.; Hu, D.; and Feng, J. 2020. Do We Really Need to Access the Source Data? Source Hypothesis Transfer for Unsupervised Domain Adaptation. In *ICML*, volume 119 of *Proceedings of Machine Learning Research*, 6028–6039. PMLR.

Liang, J.; Hu, D.; Feng, J.; and He, R. 2022a. DINE: Domain Adaptation from Single and Multiple Black-box Predictors. In *CVPR*, 7993–8003. IEEE.

Liang, J.; Hu, D.; Wang, Y.; He, R.; and Feng, J. 2022b. Source Data-Absent Unsupervised Domain Adaptation Through Hypothesis Transfer and Labeling Transfer. *TPAMI*, 44(11): 8602–8617.

Litrico, M.; Del Bue, A.; and Morerio, P. 2023. Guiding Pseudo-labels with Uncertainty Estimation for Test-Time Adaptation. *arXiv preprint arXiv:2303.03770*.

Liu, Y.; Yeh, Y.; Fu, T.; Wang, S.; Chiu, W.; and Wang, Y. F. 2018. Detach and Adapt: Learning Cross-Domain Disentangled Deep Representation. In *CVPR*, 8867–8876. Computer Vision Foundation / IEEE Computer Society.

Long, M.; Zhu, H.; Wang, J.; and Jordan, M. I. 2016. Unsupervised Domain Adaptation with Residual Transfer Networks. In *NeurIPS*, 136–144.

Long, M.; Zhu, H.; Wang, J.; and Jordan, M. I. 2017. Deep Transfer Learning with Joint Adaptation Networks. In *ICML*, volume 70 of *Proceedings of Machine Learning Research*, 2208–2217. PMLR.

Moon, S.; and Carbonell, J. G. 2017. Completely Heterogeneous Transfer Learning with Attention - What And What Not To Transfer. In *IJCAI*, 2508–2514. ijcai.org.

Peng, X.; Bai, Q.; Xia, X.; Huang, Z.; Saenko, K.; and Wang, B. 2019. Moment Matching for Multi-Source Domain Adaptation. In *ICCV*, 1406–1415. IEEE.

Peng, X.; Usman, B.; Kaushik, N.; Hoffman, J.; Wang, D.; and Saenko, K. 2017. VisDA: The Visual Domain Adaptation Challenge. *CoRR*, abs/1710.06924.

Perone, C. S.; Ballester, P. L.; Barros, R. C.; and Cohen-Adad, J. 2019. Unsupervised domain adaptation for medical imaging segmentation with self-ensembling. *NeuroImage*, 194: 1–11.

Saenko, K.; Kulis, B.; Fritz, M.; and Darrell, T. 2010. Adapting Visual Category Models to New Domains. In *ECCV*, volume 6314 of *Lecture Notes in Computer Science*, 213–226. Springer.

Saito, K.; Ushiku, Y.; and Harada, T. 2017. Asymmetric Tri-training for Unsupervised Domain Adaptation. In *ICML*,

volume 70 of *Proceedings of Machine Learning Research*, 2988–2997. PMLR.

Saito, K.; Watanabe, K.; Ushiku, Y.; and Harada, T. 2018. Maximum Classifier Discrepancy for Unsupervised Domain Adaptation. In *CVPR*, 3723–3732. Computer Vision Foundation / IEEE Computer Society.

Singh, A. 2021. CLDA: Contrastive Learning for Semi-Supervised Domain Adaptation. In *NeurIPS*, 5089–5101.

Sivaprasad, P. T.; and Fleuret, F. 2021. Uncertainty Reduction for Model Adaptation in Semantic Segmentation. In *CVPR*. Computer Vision Foundation / IEEE.

Sohn, K.; Berthelot, D.; Carlini, N.; Zhang, Z.; Zhang, H.; Raffel, C.; Cubuk, E. D.; Kurakin, A.; and Li, C. 2020. FixMatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. In *NeurIPS*.

Sun, B.; and Saenko, K. 2016. Deep CORAL: Correlation Alignment for Deep Domain Adaptation. In *ECCV Workshops*, volume 9915 of *Lecture Notes in Computer Science*, 443–450.

Sun, S.; Zhang, B.; Xie, L.; and Zhang, Y. 2017. An unsupervised deep domain adaptation approach for robust speech recognition. *Neurocomputing*, 257: 79–87.

Thota, M.; and Leontidis, G. 2021. Contrastive Domain Adaptation. *CoRR*, abs/2103.15566.

Tian, Q.; Ma, C.; Zhang, F.; Peng, S.; and Xue, H. 2021. Source-Free Unsupervised Domain Adaptation with Sample Transport Learning. *J. Comput. Sci. Technol.*, 36(3): 606–616.

Tzeng, E.; Hoffman, J.; Darrell, T.; and Saenko, K. 2015. Simultaneous Deep Transfer Across Domains and Tasks. In *ICCV*, 4068–4076. IEEE Computer Society.

Tzeng, E.; Hoffman, J.; Saenko, K.; and Darrell, T. 2017. Adversarial Discriminative Domain Adaptation. In *CVPR*, 2962–2971. IEEE Computer Society.

Tzeng, E.; Hoffman, J.; Zhang, N.; Saenko, K.; and Darrell, T. 2014. Deep Domain Confusion: Maximizing for Domain Invariance. *CoRR*, abs/1412.3474.

van den Oord, A.; Li, Y.; and Vinyals, O. 2018. Representation Learning with Contrastive Predictive Coding. *CoRR*, abs/1807.03748.

Venkateswara, H.; Eusebio, J.; Chakraborty, S.; and Panchanathan, S. 2017. Deep Hashing Network for Unsupervised Domain Adaptation. In *CVPR*, 5385–5394. IEEE Computer Society.

Vincent, P.; Larochelle, H.; Bengio, Y.; and Manzagol, P. 2008. Extracting and composing robust features with denoising autoencoders. In *ICML*, volume 307 of *ACM International Conference Proceeding Series*, 1096–1103. ACM.

Wang, H.; Xiao, R.; Li, Y.; Feng, L.; Niu, G.; Chen, G.; and Zhao, J. 2022. PiCO: Contrastive Label Disambiguation for Partial Label Learning. In *ICLR*. OpenReview.net.

Wang, R.; Wu, Z.; Weng, Z.; Chen, J.; Qi, G.; and Jiang, Y. 2023. Cross-Domain Contrastive Learning for Unsupervised Domain Adaptation. *TMM*, 25: 1665–1673.

Wei, T.; Shi, J.; Li, Y.; and Zhang, M. 2022. Prototypical Classifier for Robust Class-Imbalanced Learning. In *PAKDD*, volume 13281 of *Lecture Notes in Computer Science*, 44–57. Springer.

Wilson, G.; and Cook, D. J. 2020. A Survey of Unsupervised Deep Domain Adaptation. *ACM Trans. Intell. Syst. Technol.*, 11(5): 51:1–51:46.

Wu, Z.; Wei, T.; Jiang, J.; Mao, C.; Tang, M.; and Li, Y. 2021. NGC: A Unified Framework for Learning with Open-World Noisy Data. In *ICCV*, 62–71. IEEE.

Xiao, Z.; Wang, H.; Jin, Y.; Feng, L.; Chen, G.; Huang, F.; and Zhao, J. 2023. SPA: A Graph Spectral Alignment Perspective for Domain Adaptation. *CoRR*, abs/2310.17594.

Xie, S.; Zheng, Z.; Chen, L.; and Chen, C. 2018. Learning Semantic Representations for Unsupervised Domain Adaptation. In *ICML*, volume 80 of *Proceedings of Machine Learning Research*, 5419–5428. PMLR.

Yang, C.; Guo, X.; Chen, Z.; and Yuan, Y. 2022. Source free domain adaptation for medical image segmentation with fourier style mining. *Medical Image Anal.*, 79: 102457.

Yang, J.; Peng, X.; Wang, K.; Zhu, Z.; Feng, J.; Xie, L.; and You, Y. 2023. Divide to Adapt: Mitigating Confirmation Bias for Domain Adaptation of Black-Box Predictors. In *ICLR*. OpenReview.net.

Yang, S.; Wang, Y.; van de Weijer, J.; Herranz, L.; and Jui, S. 2021. Exploiting the Intrinsic Neighborhood Structure for Source-free Domain Adaptation. In *NeurIPS*, 29393–29405.

Ye, B.; Gan, K.; Wei, T.; and Zhang, M. 2023. Bridging the Gap: Learning Pace Synchronization for Open-World Semi-Supervised Learning. *CoRR*, abs/2309.11930.

Zhang, C.; Bengio, S.; Hardt, M.; Recht, B.; and Vinyals, O. 2017. Understanding deep learning requires rethinking generalization. In *ICLR*. OpenReview.net.

Zhang, H.; Cissé, M.; Dauphin, Y. N.; and Lopez-Paz, D. 2018. mixup: Beyond Empirical Risk Minimization. In *ICLR*. OpenReview.net.

Zhang, H.; Zhang, Y.; Jia, K.; and Zhang, L. 2021. Unsupervised Domain Adaptation of Black-Box Source Models. In *BMVC*, 147. BMVA Press.

Zhuang, F.; Cheng, X.; Luo, P.; Pan, S. J.; and He, Q. 2018. Supervised Representation Learning with Double Encoding-Layer Autoencoder for Transfer Learning. *ACM Trans. Intell. Syst. Technol.*, 9(2): 16:1–16:17.

Zhuo, J.; Wang, S.; Zhang, W.; and Huang, Q. 2017. Deep Unsupervised Convolutional Domain Adaptation. In *Proceedings of the 2017 ACM on Multimedia Conference, MM 2017, Mountain View, CA, USA, October 23-27, 2017*, 261–269. ACM.