ME-SFDA: Marginal Exploration with Pyramidal Atkinson-Shiffrin Memory for Source-Free Domain Adaptation

Chunzhi Liu[1], **Yuwu Lu**[1,*]

[1]School of Artificial Intelligence, South China Normal University, China
2023024323@m.scnu.edu.cn, luyuwu2008@163.com

Abstract

Source-free domain adaptation (SFDA) aims to transfer knowledge from a source domain to an unlabeled target domain without requiring access to source data. Although previous works have focused on clustering target domain samples from continuous training, there are still some challenges: i) More source domain knowledge is forgotten with more training epochs. ii) Achieving better learning results often requires increased computational resources. To solve these problems, we propose a novel Marginal Exploration for Source-Free Domain Adaptation (ME-SFDA) method, which is a multi-scale information fusion learning based on our designed Pyramidal Atkinson-Shiffrin memory. Specifically, we design a two-step module to split samples into clustered cores and response scatters by sensory memory. Then, a novel technique is proposed for clustering samples in a hierarchical way, utilizing long-term memory to cluster cores derived from splitting the samples earlier and guide response scatters. To effectively divide samples of different classes, we propose a method that encourages unambiguous cluster assignments for the samples using multi-scale fusion information. To verify the generality of our approach, we not only discuss the UDA and SFDA tasks but also apply it to the semi-supervised domain adaptation (SSDA), which utilizes a few labeled target samples based on UDA. Extensive experiments on all utilized standard benchmarks indicate that our approach outperforms previous SOTA methods.

Code — https://github.com/ushengvv/ME-SFDA

Introduction

Deep neural networks (DNNs) have achieved several breakthroughs in visual recognition tasks (Liu et al. 2022; Li et al. 2022; Zhou et al. 2022). However, the performance of DNNs models is degraded when they are applied to domain shift (Gretton et al. 2009; Xia and Ding 2020) with an unseen target domain. To address this problem, various unsupervised domain adaptation (UDA) methods (Huang et al. 2022; Wu et al. 2023; Wang et al. 2023a,b) have been proposed. Most UDA techniques require access to unlabeled target domain data and labeled source domain data during adaptation, aiming to address the distribution gap and decrease the influence of domain shift. However, the application of UDA techniques is limited in some real scenarios with private source data or devices with limited computing power.

Due to the absence of a requirement to access source samples during adaptation, source-free domain adaptation (SFDA) (Liang, Hu, and Feng 2020; Yang et al. 2022; Litrico et al. 2023; Karim et al. 2023; Chen et al. 2022; Yang et al. 2021c; Hwang et al. 2024) has received widespread attention, as it can effectively complete tasks under this limitation. SFDA requires only unlabeled target domain data and a pretrained source model available. Most of the SFDA methods used in previous works have focused on self-training (Liang, Hu, and Feng 2020; Karim et al. 2023), self-regularization (Yang et al. 2022, 2021c), contrastive learning (Yang et al. 2022; Karim et al. 2023; Chen et al. 2022), and auxiliary self-supervision (Litrico et al. 2023; Chen et al. 2022). Despite the contributions of previous SFDA methods, they neglected the preservation of source class knowledge, leading to rapid degradation of model ability to distinguish similar classes after training.

To address this challenge, we draw inspiration from the Atkinson-Shiffrin memory model (Atkinson and Shiffrin 1968) in neuroscience, which mimics human brain learning by dividing memory into three types: sensory, short-term, and long-term memory. The sensory memory stores information for a very short duration (*e.g.*, 0.5–1.0 seconds) and then decays. The short-term memory stores selected information and usually decays after a longer duration (*e.g.*, 18–20 seconds). The long-term memory stores information from the short-term memory and retains it almost permanently. In computer vision, the human memory process has attracted the attention of several works (Zhang et al. 2023; Cheng and Schwing 2022). However, these methods are limited to storing and refining the feature knowledge extracted by the model without jointly learning with other knowledge, which leads to the enlightenment of Atkinson-Shiffrin memory to the model is not fully utilized.

In this work, we propose a novel SFDA approach called ME-SFDA based on our designed Pyramidal Atkinson-Shiffrin memory, which enhances similar knowledge aggregation and improves learning efficiency. Pyramidal Atkinson-Shiffrin memory is designed to store multi-scale fusion information in a pyramid-shaped structure based on its storage space, which contains three types of memory

*Corresponding author.

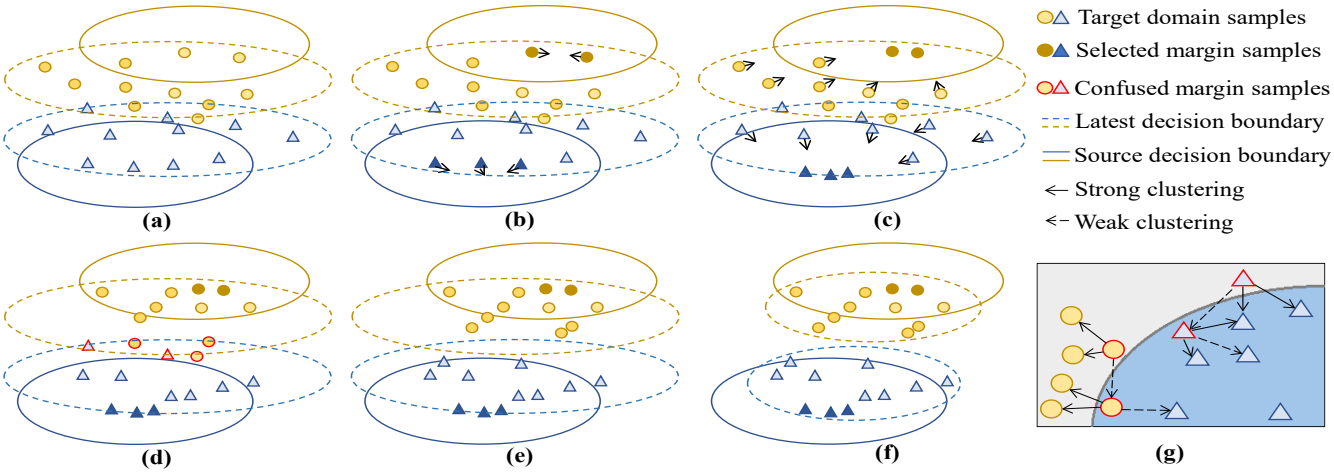


Figure 1: An example to illustrate ME-SFDA. We first set up a source discriminator that will not update the weight and a dynamic target discriminator, as shown in (a). In (b) and (c), our two-step mechanism initially splits the samples into two groups via the source discriminator, with selected margin samples employing a cohesion strategy and the remaining samples employing the memorized feature clustering toward the selected margin samples. In (d), the two-step mechanism further splits the non-selected margin samples into two groups and introduces the memory difference separation for the confused margin samples in (g), and the result in (e) shows that the samples in the confusion domain are perfectly separated. Finally, our clustering loss updates the target discriminator to facilitate cross-domain cluster-wise feature alignment, as shown in (f). In these process, the Pyramidal Atkinson-Shiffrin memory is used to store samples related multi-scale fusion information.

(sensory memory, short-term memory, and long-term memory), interacting in a hybrid manner through forward memorization and backward calibration. Based on the proposed Pyramidal Atkinson-Shiffrin memory, we exploit similarities between marginal samples at the decision boundary to cluster similar samples and minimize their discrepancies, while designing distinct strategies for two diagonal boundary samples. In our approach, we first use the sensory memory to split samples into clustered cores and response scatters and then design the memorized feature clustering to attract the response scatters toward the clustered cores. Moreover, we set up a label aggregation mechanism that makes the clustered cores more compact using the long-term memory. Second, we split the response scatters into confused scatters and neighbor scatters. Confused scatters generally do not change much in the clustering strategy because they are usually far from the clustered cores of two different domains. To correctly classify the confused scatters, we propose the memory difference separation based on our multi-scale information fusion. Finally, a contrastive learning strategy is designed to contrast the short-term memory of the predicted samples, which avoids model degeneracy and encourages the samples to have unambiguous cluster assignments. Different from other memory-based methods (Yang et al. 2022, 2021c; Chen et al. 2022), ME-SFDA subdivides memories and focuses learning on important memories, which improves the model's discrimination with less consumption. ME-SFDA is the first to explore the memory process by performing a pyramidal subdivision of inter-sample relationships and conducting multi-scale information fusion learning. As shown in Figure 1, we explain how our proposed approach avoids the problem of source knowledge being forgotten. To verify the generality of our method, we also apply ME-SFDA to the semi-supervised domain adaptation (SSDA) problem, which has recently become a popular problem that transfers source knowledge to the target domain with a few labeled samples during adaptation.

Related Work

Source-Free Domain Adaptation. SFDA aims to adapt unlabeled target domain samples only using the pretrained model without source domain samples. Previous state-of-the-art works (Liang, Hu, and Feng 2020; Ding et al. 2022; Karim et al. 2023; Chen et al. 2022; Yang et al. 2021c) mostly guide the source pretrained model to generate pseudolabels via a self-training mechanism. The early SFDA work (Liang, Hu, and Feng 2020) utilized the knowledge of per-class cluster structure with self-training in the feature space to refine the pseudolabels, but the label information formulated based on cluster structure is severely misleading or noisy. To reduce the impact of these negative factors, NRC (Yang et al. 2021c) observes that target data can form clear clusters without aligning with the source domain classifier, proposing a self-regularization method that allows each sample to attract neighboring samples with high affinity. Although samples of each class are clustered more compactly in the NRC (Yang et al. 2021c), indiscriminately attracting neighboring points causes a small number of samples on the edge of the decision boundary sticking to each other. This causes the formation of some independent and confused clusters, leading to the discarding of misleading samples and partial source knowledge. In the last two years, several methods have been proposed, which do not require the discarding of indistinguishable samples. AaD

(Yang et al. 2022) combines NRC (Yang et al. 2021c) and contrastive learning to optimize the objective of prediction consistency. GPUE (Litrico et al. 2023) proposes a negative pair exclusion strategy to refine pseudolabels by aggregating knowledge from neighboring samples. AdaContrast (Chen et al. 2022) applies contrastive learning jointly with pseudolabeling to contrast the construction of negative and positive pairs by employing a memory bank. The bank is created to distribute cluster knowledge and refine pseudolabels via soft voting among neighboring samples. C-SFDA (Karim et al. 2023) employs a curriculum learning scheme to prevent label noise propagation during adaptation stages by the selection of high-reliability samples. However, these methods require costly storage space or training time to make different samples more differentiated.

Semi-Supervised Domain Adaptation. Due to the availability of a few labeled target samples, SSDA comes with only a small cost for obtaining sample labels to significantly improve the classification performance on the target domain compared to UDA (Wang et al. 2023a; Yang et al. 2021a; Melas-Kyriazi and Manrai 2021; Gao et al. 2023; Zhou et al. 2023). Previous SSDA methods can be roughly divided into two types that differ in how they cluster assignments and align interdomain feature distributions. One is adversarial training-based methods, such as in (Singh 2021; Yang et al. 2021b; Li et al. 2021b; Huang, Zhu, and Chen 2023; Yan et al. 2022); the second is maximum mean discrepancy (MMD)-based methods, such as in (Ngo et al. 2023; Li, Li, and Yu 2023; Yu and Lin 2023; Zhu et al. 2020). Among the MMD methods, G-ABC (Li, Li, and Yu 2023) proposes adaptive betweenness clustering to facilitate interdomain semantic alignment. Compared to MMD methods, adversarial training-based methods are more popular. For example, Clda (Singh 2021) employs contrastive learning and pseudolabeling to reduce interdomain and intradomain discrepancies, and ECACL (Li et al. 2021b) proposes a framework that integrates multiple mutually complementary domain alignment techniques through sample confrontation. Some prototype-based methods based on adversarial training have emerged in recent years. MCL (Yan et al. 2022) designs a prototype-based optimal transport method to regularize the consistency of different views at the interdomain, inner domain, and sample levels. Based on MCL, ProML (Huang, Zhu, and Chen 2023) introduces a pseudolabel aggregation approach to align the prototype and feature distribution in the target domain.

Proposed Method

An overview of ME-SFDA and the entire process from the pretraining period to the completed training is shown in Figure 2. We start by defining the unlabeled target domain dataset as $D_t = \{x_t^i\}_{i=1}^N$, which contains N samples. Then, we parameterize our model by θ, which is composed of two components, namely, a feature extractor and a classifier. The feature extractor F is denoted by $f_\theta : x^i \to z^i \in \mathbb{R}^k$, where k is the dimension of the feature space and z is denoted as a k-dimensional feature. And the classifier C is denoted by $c_\theta : z^i \to y^i \in \mathbb{R}^n$, where n represents the number of predicted classes and y is the prediction output.

Pyramidal Atkinson-Shiffrin Memory with Two-Step Strategy

We describe the process of two-step strategy first, as it is a vital part of our proposed clustering learning strategy. At the beginning of our training period, we introduce the source classifier and set it as two classifiers with the same weight, the fixed-weight classifier C_{fix} and target classifier C, according to whether the weight changes in the subsequent process. In the first step of segmentation, the predictions of C_{fix} and C are stored in the sensory memory buffer. This buffer is used to split all samples into different types in two steps. Our first set of division formulas is defined as follows:

$$p_1(x) = C_{fix}(f_\theta(x)),\ \sigma_1(x) = softmax(p_1(x)), \quad (1)$$

$$K(p) = TopK(p, 0) - TopK(p, 1), \quad (2)$$

$$\tau_{p1} = \frac{\lambda_1}{N}\sum_{i=1}^{N} K\left(p_1^i\right),\ \tau_{\sigma 1} = \frac{\lambda_1}{N}\sum_{i=1}^{N} TopK\left(\sigma_1^i, 0\right), \quad (3)$$

$$Category1_i = \begin{cases} 1,\ if\ K(p_1^i) > \tau_{p1}\ and\ TopK(\sigma_1^i, 0) > \tau_{\sigma 1} \\ 0,\ otherwise \end{cases}, \quad (4)$$

where $TopK(\cdot, k)$ denotes the first-k value of the predicted classification output and λ_1 is the first hyperparameter of our proposed segmentation, which determines the segmentation ratio in the first step. When $Category1_i = 1$, the i-th sample is assigned to a clustered core; otherwise, it is a response scatter. In the first stage, we combine the feature extractor f_θ that has learned new knowledge with C_{fix} that contains source knowledge. Due to the divergence between the two classifiers, and under the effect of double filtering, we can identify a small but highly reliable set of clustered cores that are recognized by both the source and target domains.

Then, we choose those samples where $Category1_i = 0$ and further split them into two categories using the target classifier, *i.e.*, confused scatters and neighbor scatters. Our second set of division formulas is defined as follows:

$$p_2(x) = C(f_\theta(x)), \quad (5)$$

$$\tau_{p2} = \frac{\lambda_2}{N}\sum_{i=1}^{N} K\left(p_2^i\right), \quad (6)$$

$$Category2_i = \begin{cases} 1, & if\ K(p_2^i) > \tau_{p2} \\ 0, & otherwise \end{cases}, \quad (7)$$

where λ_2 represents a hyperparameter used to determine the segmentation ratio in the second step. When $Category2_i = 1$, the i-th sample will be assigned to a neighbor scatter; otherwise, it will be a confused scatter. In the second step, instead of performing detailed filtering as in Eq. (3), we only need to determine the decision boundary according to the distribution of target samples using Eq. (6). Since C continues to learn the target knowledge, it inevitably generates prediction divergence with C_{fix}. λ_1 and λ_2 are used to control the divergence ratio, the former determines the number of clustered cores, and the latter controls the decision boundary.

The significance of setting up two classifiers is that they learn target domain knowledge while maximizing the use

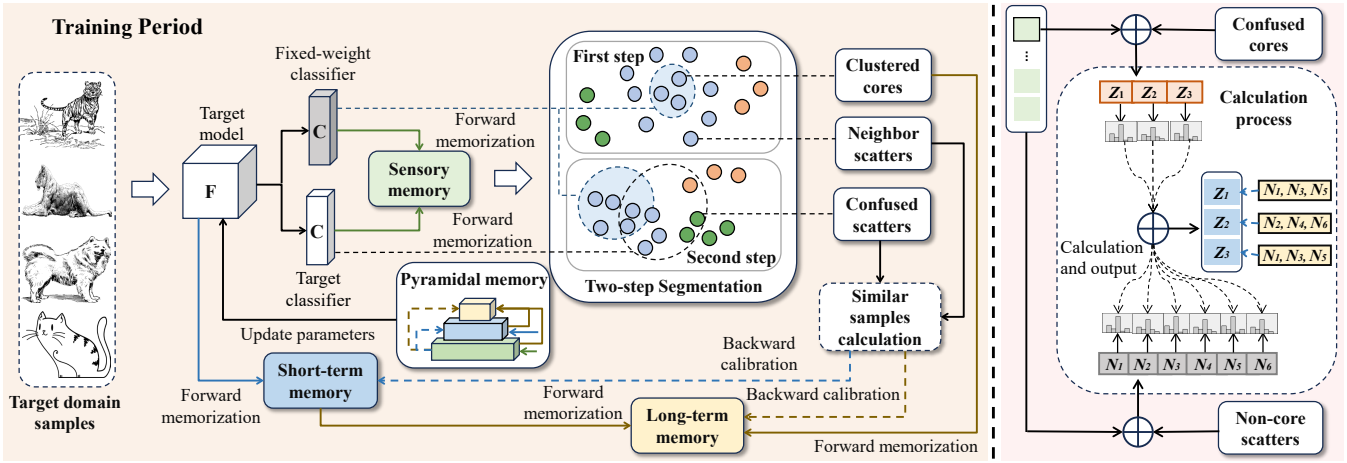


Figure 2: Overview of the completed training period of ME-SFDA. In the initial training, we obtain a source classifier from the pretraining period and set it to a fixed-weight classifier and a dynamic target classifier. The fixed-weight classifier and target classifier output the multi-scale information to the sensory buffer and use the sensory memory to split target domain samples into clustered cores, neighbor scatters, and confused scatters. Then, the short-term buffer is calibrated by calculating and selecting samples from these three categories, which stores non-core sample information. As shown in right, the calculation module receives the input of confused samples and looks for similar non-core samples from the sensory memory buffer to perform pairing. The long-term memory buffer permanently stores the multi-scale information of the cluster cores during training, and this information is continuously calibrated and updated through the sensory and short-term memories.

of the classification knowledge of the source model. In our research, by comparing the sample predictions stored in the sensory memory buffer, we can more effectively exploit their connections and differences to facilitate both intradomain semantic alignment and interdomain semantic segmentation. The entire process of two-step strategy is completed in Pyramid Atkinson-Shiffrin memory. For the memory update, we adopt a hybrid manner including the forward update and the backward calibration. For the sensory memory, it only forward updates to store sample information directly output from the model; for the short-term memory, it forward updates response scatters information, and then backward calibrates the selection of response scatters by the sensory memory; and for the long-term memory, it maintains an additional index set of low cost to store sample associations, which forward updates clustered cores information from the sensory and short-term memories, and the selection of core samples is backward calibrated by the sensory memory.

Memorized Feature Clustering

In the human knowledge learning process (Atkinson and Shiffrin 1968), clustered association of similar knowledge is an essential learning step. Therefore, we propose a clustering strategy to centralize clustered cores and cluster response scatters based on the sample spatial distances. We first simply centralize clustered cores stored in the long-term memory buffer to determine the cluster center, using standard cross-entropy loss as follows:

$$\mathcal{L}_{ce} = -\frac{1}{B}\sum_{i=1}^{B} Y^i \log p(x^i), \tag{8}$$

$$p\left(x\right) = C(f_\theta\left(x\right)), \tag{9}$$

where B is the set of clustered cores in the same batch; Y^i represents the predicted pseudolabel of the i-th core sample; and the predicted most likely class is used as their pseudolabels. This cross-entropy loss quickly aggregates the core of each class to greatly reduce subsequent clustering costs.

Then, we utilize the latest clustered core predictions stored in sensory buffer to guide response scatters stored in short-term buffer in clustering. The computations in this process are computed as follows:

$$S_R = [\ \sigma(p(x_R^1)),\ \sigma(p(x_R^2)), ...,\ \sigma(p(x_R^i))\ ]^T,\ i \leq N_R, \tag{10}$$

$$\mathcal{L}_{dis} = \frac{1}{R}\sum_{i=1}^{R} D_{KL}(S_R^i\ ||\ \sum_{j=1}^{B}\sigma(p(x^j))), \tag{11}$$

where $\sigma(\cdot)$ denotes the $softmax$ function; $D_{KL}(\cdot)$ is the Kullback-Leibler divergence; R represents the set of response scatters in a mini-batch; and S_R is the classification prediction in the short-term memory, containing N_R response scatters.

Because $\mathcal{L}_{ce}$ and $\mathcal{L}_{dis}$ are two stages of clustering, they are both designed to minimize the discrepancies among samples with similar features, and we combine these losses as follows:

$$\mathcal{L}_{clu} = \mathcal{L}_{dis} + \alpha\mathcal{L}_{ce}\ , \tag{12}$$

where α is a hyperparameter used to adjust the offset of similar samples.

Multi-scale Memory Difference Separation

Our proposed clustering strategy can cluster feature similar samples quickly and efficiently, but its influence on some

samples in the confusion area is still limited. In this section, we address this issue from the perspective of sample refinement. To correctly classify the samples in the confusion area, we propose a boundary sample separation strategy based on multi-scale information fusion. In this section, we utilize previously confused scatters stored in the short-term memory to focus on solving the problem of samples in the confusion region. We initially set S as a classification part in sensory buffer and G as a feature part; each iteration only updates θ of the current batch S and G, which can be defined as follows:

$$G = [\ f_\theta(x^1),\ f_\theta(x^2), ...,\ f_\theta(x^i)\ ]^T,\ i \leq N, \tag{13}$$

$$S = [\ \sigma(p(x^1)),\ \sigma(p(x^2)), ...,\ \sigma(p(x^i))\ ]^T,\ i \leq N. \tag{14}$$

Then, we utilize the pairwise feature similarity pseudolabels of the response scatter pairs in a mini-batch and establish pairwise attractions between confused scatters and others using the short-term buffer, which are computed as follows:

$$S_F = [\ \sigma(p(x_F^1)),\ \sigma(p(x_F^2)), ...,\ \sigma(p(x_F^i))\ ]^T,\ i \leq N_F, \tag{15}$$

$$\mathcal{L}_{sep} = \frac{1}{F}\sum_{i=1}^{F} D_{KL}(S_F^i\ ||\ S^{sim(G_F^i, G, \phi)}), \tag{16}$$

where F is the part of confused scatters in short-term buffer; S_F is the confused scatter classification prediction of the current batch, containing N_F response scatters; G_F is the confused scatter feature as a part of short-term buffer; $sim(\alpha, \beta, \phi)$ is the cosine similarity between α and β, and its output is the index of the first ϕ samples similar to α, where these indexes calibrate the long-term memory after the calculation is complete; and ϕ denotes a hyperparameter to determine whether the confused scatters are influenced by the number of neighbors. These confused scatters can be attracted by neighbors, and their attractiveness is determined by feature similarity.

To address the overfitting of the model, we propose a contrastive learning approach to increase the prediction diversity and discriminability, which can be expressed as:

$$\mathcal{L}_{cl} = \frac{1}{R}\sum_{i=1}^{R}\sum_{j=1}^{R} P_i^T P_j \cdot \left(\mathbf{I} - \frac{tr\{P_i^T P_j\}}{\|P_i\| \cdot \|P_j\|}\right), \tag{17}$$

where P represents the classification prediction of samples in a mini-batch; $\mathbf{I}$ is a unit vector; and $tr\{\cdot\}$ is the trace of a matrix. The utilization of entropy minimization in Eq. (8) may result in sustained convergence towards the predominant categories, consequently narrowing the predictive diversity for the less frequent categories (Cui et al. 2020). Therefore, it is essential to limit this impact while ensuring effective aggregation in the feature space using entropy minimization. In our method, such contrastive learning promotes a clear distinction between different clusters by driving the predictions of target features away from each other and buffering the attraction efficiency of clustered cores on response scatters, which significantly avoids model prediction degradation.

Finally, by combining $\mathcal{L}_{clu}$, $\mathcal{L}_{sep}$, and $\mathcal{L}_{cl}$, the total loss of ME-SFDA is as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{clu} + \mathcal{L}_{sep} + \mathcal{L}_{cl}. \tag{18}$$

Confirming Generalization

To verify the generality of our ME-SFDA approach, we also extend it to the SSDA problem in our work. In SSDA, we can obtain a few labeled target samples during the training period, and they can help the model transfer knowledge representations from the source domain to the target domain. We define the labeled target domain dataset as $D_l = \{(x_l^i, y_l^i)\}_{i=1}^{N_l}$, which contains N_l samples, and we only make use of them during the training period. Moreover, a small modification was made so that we can effectively utilize these labeled samples; that is, let $B^* \subseteq B \cup D_l$ replace B in Eqs. (8) and (11). We add the labeled samples to each mini-batch as evenly as possible during data input. When other loss functions remain unchanged, this operation helps concentrate the clustered cores in more appropriate locations, facilitating the easier clustering of response scatters with similar features. In the experiments, we called this SSDA method as ME-SSDA.

Experiments

Datasets

Our experiments are performed on four datasets, including **Office** (Saenko et al. 2010), **Office-Home** (Venkateswara et al. 2017), and **VisDA** (Peng et al. 2017). The **Office** dataset is a common domain adaptation dataset that consists of 3 domains with 31 categories: Dslr (D), Webcam (W), and Amazon (A). In the experiments, all 6 pairs of domain scenarios were completed. **Office-Home** is a widely used domain adaptation dataset with 4 domains from 65 classes: Real (R), Clipart (C), Art (A), and Product (P). There are 12 adaptation scenarios with different scales of domain gaps to overcome. **VisDA** is a large benchmark dataset for domain adaptation tasks and presents a greater challenge with 12 categories, ranging from synthetic to real.

Baselines and Results

We choose several related methods for comparison with our methods, including SHOT (Liang, Hu, and Feng 2020), SFDA-DE (Ding et al. 2022), AaD (Yang et al. 2022), GPUE (Litrico et al. 2023), C-SFDA (Karim et al. 2023), AdaContrast (Chen et al. 2022), NRC (Yang et al. 2021c), CDAC (Li et al. 2021a), S3D (Yoon, Kang, and Cho 2022), ProML (Huang, Zhu, and Chen 2023), MCL (Yan et al. 2022), NBF (Song et al. 2024), DaC (Zhang et al. 2022), NRC+ELR (Yi et al. 2023), ASM (Jing et al. 2024), SF(DA)^2 (Hwang et al. 2024), and ICPR (Tian et al. 2025). For a fair comparison, all baseline results were obtained from the original papers, their associated codebases, or follow-up work. As shown in Tables 1, 2, and 3, our approach achieves SOTA performance in all benchmarks, demonstrating the effectiveness of marginal exploration with the Pyramidal Atkinson-Shiffrin memory in enhancing predictive discrimination under SFDA and SSDA.

Method	Publication	SF/SS	A→C	A→P	A→R	C→A	C→P	C→R	P→A	P→C	P→R	R→A	R→C	R→P	Mean
SHOT	*ICML'20*	SF	57.1	78.1	81.5	68.0	78.2	78.1	67.4	54.9	82.2	73.3	58.8	84.3	71.8
DaC	*NIPS'22*	SF	59.1	79.5	81.2	69.3	78.9	79.2	67.4	56.4	82.4	74.0	61.4	84.4	72.8
AaD	*NIPS'22*	SF	59.3	79.3	82.1	68.9	79.8	79.5	67.2	57.4	83.1	72.1	58.5	85.4	72.7
SFDA-DE	*CVPR'22*	SF	59.7	79.5	82.4	69.7	78.6	79.2	66.1	57.2	82.6	73.9	60.8	85.5	72.9
C-SFDA	*CVPR'23*	SF	60.3	80.2	82.9	69.3	80.1	78.8	67.3	58.1	**83.4**	73.6	**61.3**	86.3	73.5
NRC+ELR	*ICLR'23*	SF	58.4	78.7	81.5	69.2	79.5	79.3	66.3	58.0	82.6	73.4	59.8	85.1	72.6
ASM	*TIP'24*	SF	56.9	79.1	82.9	69.5	79.6	79.7	67.9	55.1	82.6	74.7	60.5	84.8	72.8
ICPR	*TNNLS'25*	SF	60.1	79.4	80.7	**69.9**	79.6	79.9	**69.8**	58.6	81.7	73.8	59.4	84.9	73.1
ME-SFDA	*Our*	SF	**61.3**	**80.8**	**82.9**	69.2	**81.3**	**80.8**	66.7	**58.9**	83.1	**74.8**	60.9	**86.7**	**74.0**
CDAC (1/3)	*CVPR'21*	SS	61.7/66.0	76.0/80.7	78.4/80.7	64.8/67.5	75.4/81.4	75.7/80.3	64.9/67.6	60.6/67.0	81.0/81.9	72.9/72.3	61.9/67.9	83.6/85.9	71.4/74.9
S3D (1/3)	*WACV'22*	SS	58.3/61.5	76.1/76.9	79.8/79.6	65.2/66.3	73.9/76.3	74.3/74.8	64.7/66.2	57.5/58.2	81.9/82.4	72.6/71.8	60.2/63.6	82.2/82.8	70.6/71.7
MCL (1/3)	*IJCAI'22*	SS	64.4/**67.9**	79.6/83.7	81.5/82.7	69.0/71.7	79.7/84.6	78.8/81.9	68.4/70.2	61.3/68.1	81.9/83.4	74.1/76.3	66.8/70.0	86.0/**88.9**	74.3/77.5
ProML (1/3)	*IJCAI'23*	SS	**64.7/67.9**	79.9/84.2	81.9/82.4	69.2/72.3	80.8/84.2	79.3/82.8	**69.3**/72.8	**61.7/68.9**	82.0/83.9	73.9/76.9	**67.8/72.1**	86.3/88.7	74.7/78.1
NBF (1/3)	*ECCV'24*	SS	58.7/62.2	77.6/81.0	79.6/79.7	67.6/68.8	75.1/85.4	75.4/78.6	66.2/67.7	60.1/61.7	79.9/79.5	68.1/69.0	62.6/64.1	83.3/88.2	71.2/73.8
ME-SSDA (1/3)	*Ours*	SS	62.3/66.5	**83.1/84.5**	**83.2/84.7**	**71.9/74.7**	**82.8/85.1**	**82.0/85.0**	68.8/**73.0**	61.3/66.9	**83.3/84.6**	**76.5/79.9**	62.6/67.1	**86.9**/88.6	**75.4/78.5**

Table 1: Performance under SFDA (SF) and SSDA (SS) settings on the **Office-Home** dataset using the ResNet-50 backbone. The top-performing results in their corresponding fields are highlighted in **bold**. (1/3) represents the 1-shot/3-shot labeled target domain samples used in SSDA.

Method	Publication	SF/SS	A→D	A→W	D→W	W→D	D→A	W→A	Mean
SHOT	*ICML'20*	SF	93.1	90.9	98.8	99.9	74.5	74.8	88.7
NRC	*NIPS'21*	SF	96.0	90.8	99.0	**100**	75.3	75.0	89.4
AaD	*NIPS'22*	SF	**96.4**	92.1	**99.1**	**100**	75.0	76.5	89.9
SFDA-DE	*CVPR'22*	SF	96.0	**94.2**	98.5	99.8	76.6	75.5	90.1
C-SFDA	*CVPR'23*	SF	96.2	93.9	98.8	99.7	77.3	77.9	90.5
NRC+ELR	*ICLR'23*	SF	93.8	93.3	98.0	**100**	76.2	76.9	89.6
SF(DA)^2	*ICLR'24*	SF	95.8	92.1	99.0	99.8	75.7	76.8	89.9
ICPR	*TNNLS'25*	SF	93.4	95.1	98.5	100	74.9	75.4	89.5
ME-SFDA	*Our*	SF	96.2	93.9	**99.1**	**100**	**77.8**	**78.6**	**91.1**
CDAC (1/3)	*CVPR'21*	SS	87.5/91.2	90.3/91.5	97.2/97.3	98.4/98.7	65.6/74.3	64.2/75.2	84.0/88.0
MCL (1/3)	*IJCAI'22*	SS	93.2/96.2	88.3/94.6	97.7/97.9	99.5/99.9	72.9/79.5	76.6/76.7	88.0/90.8
ProML (1/3)	*IJCAI'23*	SS	93.7/94.8	91.7/96.9	98.6/**99.1**	99.6/99.9	76.3/79.4	78.3/80.2	89.7/91.7
NBF (1/3)	*ECCV'24*	SS	88.3/92.3	89.2/90.9	98.1/98.5	99.4/99.8	66.4/73.2	68.3/75.4	85.9/88.4
ME-SSDA (1/3)	*Ours*	SS	**97.4/97.8**	**95.5/97.1**	**99.0**/99.0	**100/100**	**77.9/81.0**	**79.4/81.7**	**91.5/92.8**

Table 2: Performance under SFDA (SF) and SSDA (SS) settings on the **Office** dataset using the ResNet-50 backbone.

Ablation Study

We conducted extensive experiments to individually confirm the efficacy of each component of our proposed ME-SFDA approach. Table 4 shows the main ablation study results on three benchmark datasets. As demonstrated, each part of ME-SFDA contributes to improving the performance. $\mathcal{L}_{sep}$ contributes the most to improve accuracy when acting alone, with the average performance increasing from 62.3% to 75.9%. According to our analysis, $\mathcal{L}_{sep}$ excels primarily in processing boundary information, but it does not perform well in clustering similar samples independently. Therefore, this separation loss effectively handles rough clustering subtasks but proves ineffective when dealing with **Office**, which contains a considerable number of ambiguous samples. Then, we compare the loss functions as pairs. As illustrated in Table 4, the effect of combining $\mathcal{L}_{clu}$ and $\mathcal{L}_{sep}$ results in lower accuracy (66.7%) compared to other combinations of loss functions. This operation causes overfitting and leads to model degradation because $\mathcal{L}_{clu}$ causes the model to shift toward the maximum feature gap space, and $\mathcal{L}_{sep}$ delays the former in addressing boundary issues, making it difficult for the class discriminant space to shrink effectively and accurately without contrastive learning. When all losses are used, $\mathcal{L}_{cl}$ plays an auxiliary role by driving the predictions of all target features away from each other to promote a clear distinction between different clusters. For this, Figure 3 demonstrates that ME-SFDA increases prediction discriminability and diversity by adaptively adjusting the training on under-learned samples.

Parameter Analysis and Comparison

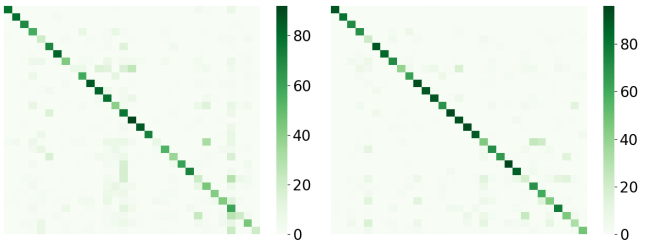
In Figure 4, we plot the trends of prediction accuracy by ME-SFDA and other approaches in the A→R subtask of the **Office-Home** dataset. As shown in Figures 4(a) and (b), the effectiveness of the proposed clustering and separation strategies to increase discriminability and robustness is demonstrated both on SSDA and SFDA. λ_1 and λ_2 are used to control the prediction divergence ratio, the former deter-

Method	Publication	plane	bike	bus	car	horse	knife	mcycle	person	plant	sktbrd	train	truck	Mean
SHOT	*ICML'20*	94.3	88.5	80.1	57.3	93.1	94.9	80.7	80.3	91.5	89.1	86.3	58.2	82.9
NRC	*NIPS'21*	96.9	89.7	84.0	59.8	95.9	96.6	86.5	80.9	92.8	92.6	90.2	60.2	85.4
AaD	*NIPS'22*	97.4	90.5	80.8	76.2	97.3	96.1	89.8	82.9	95.5	93.0	92.0	**64.7**	88.0
DaC	*NIPS'22*	96.6	86.8	86.4	78.4	96.4	96.2	**93.6**	83.8	96.8	95.1	89.6	50.0	87.3
SFDA-DE	*CVPR'22*	95.3	91.2	77.5	72.1	95.7	**97.8**	85.5	86.1	95.5	93.0	86.3	61.6	86.5
AdaContrast	*CVPR'22*	97.0	84.7	84.0	77.3	96.7	93.8	91.9	84.8	94.3	93.1	**94.1**	49.7	86.8
NRC+ELR	*ICLR'23*	97.1	89.7	82.7	62.0	96.2	97.0	87.6	81.2	93.7	94.1	90.2	58.6	85.8
C-SFDA	*CVPR'23*	97.6	88.8	86.1	72.2	97.2	94.4	92.1	84.7	93.0	90.7	93.1	63.5	87.8
SF(DA)^2	*ICLR'24*	96.8	89.3	82.9	81.4	96.8	95.7	90.4	81.3	95.5	93.7	88.5	**64.7**	88.1
ICPR	*TNNLS'25*	97.8	91.7	87.8	86.7	97.1	96.3	92.2	81.1	**97.7**	95.1	90.3	53.4	88.9
ME-SFDA	*Ours*	**98.4**	**92.6**	**90.0**	**94.2**	**97.7**	95.6	92.1	**86.8**	97.4	**95.6**	94.0	41.3	**89.6**

Table 3: Performance under SFDA settings on the **VisDA** dataset using the ResNet-101 backbone.

$\mathcal{L}_{clu}$	$\mathcal{L}_{sep}$	$\mathcal{L}_{cl}$	Office	Office-Home	VisDA	Mean
Source only			80.2	61.6	45.1	62.3
✓			85.7	64.2	74.0	74.6
	✓		84.3	63.9	79.6	75.9
		✓	84.1	56.9	51.3	64.1
✓	✓		84.8	62.5	52.9	66.7
✓		✓	89.2	73.1	82.8	81.7
	✓	✓	88.3	68.7	86.4	81.1
✓	✓	✓	**91.1**	**74.0**	**89.6**	**84.9**

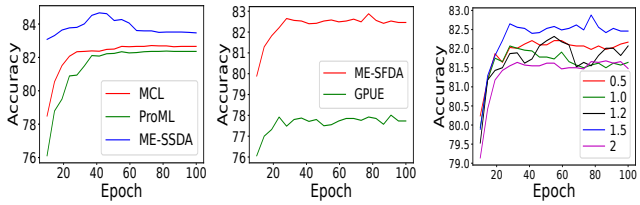
Table 4: Results (%) of Ablation Study using SFDA setting, where ResNet-50 is used for **Office**, **Office-Home**; and ResNet-101 is used for **VisDA**. ✓ denotes that the adaptation model is trained by the corresponding loss function.



(a) Source only (b) ME-SFDA

Figure 3: Classification visualization with Confusion Matrix for both the source model and ME-SFDA in the D→A subtask of the **Office** dataset.

mines the number of clustered cores, and the latter controls the segmentation decision boundary. In Figure 4(c), we discuss the impact of different values of parameter λ_1 on our model and find that a better result is obtained when λ_1 is equal to 1.5. λ_2 controls the role of $\mathcal{L}_{sep}$ by determining the ratio of confused scatters and neighbor scatters in the short-term memory buffer. As shown in Table 5, higher λ_2 is more suitable for sample-richer datasets, it plays a role in improving generalization. And in the **Office** datasets, where the samples are more concentrated, using a lower value of λ_2 can reduce the impact of confused scatters on other samples. ME-SFDA still has the potential to achieve better results.



(a) (b) (c)

Figure 4: Accuracy trends of predictions in the A→R subtask of the **Office-Home** dataset. The accuracy evolution of our proposed method and comparison methods under the SSDA (3-shot) and SFDA settings are shown in (a) and (b), respectively. The evolution of the accuracy in (c) of the predictions over epochs while varying the hyperparameter λ_1.

ME-SFDA	Office	Office-Home	VisDA	Mean
w/o Two-Step	86.5	68.8	82.9	79.4
$\lambda_2 = 0.5$	**91.1**	72.4	86.8	83.5
$\lambda_2 = 1$	90.9	72.9	88.9	84.2
$\lambda_2 = 1.6$	90.8	**73.8**	89.6	**84.7**
$\lambda_2 = 2$	89.9	**73.8**	**90.0**	84.5

Table 5: Results with different values of the hyperparameter λ_2 in Eq. (6) under the SFDA setting, when λ_1 is set to 1.

Conclusion

We propose a novel method, ME-SFDA, which is a multi-scale information fusion learning based on our Pyramidal Atkinson-Shiffrin memory. Due to the inevitable presence of some source knowledge that is forgotten with an increase in training epochs, we employ a novel two-step segmentation that splits samples into three types through the sensory memory. Then, we design multi-scale information fusion learning strategies: the memorized feature clustering guides similar samples to aggregate, thereby minimizing discrepancies among them; the memory difference separation efficiently separates these samples in the confused domain. Experiments show that ME-SFDA enhances both discriminability and robustness, while demonstrating strong generalization and achieving SOTA performance across all benchmarks.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 62176162 and in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2023A1515012875 and Grant 2022A1515140099.

References

Atkinson, R. C.; and Shiffrin, R. M. 1968. Human Memory: A Proposed System and its Control Processes. *Psychology of Learning and Motivation*, 2: 89–195.

Chen, D.; Wang, D.; Darrell, T.; and Ebrahimi, S. 2022. Contrastive test-time adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 295–305.

Cheng, H. K.; and Schwing, A. G. 2022. XMem: Long-Term Video Object Segmentation with an Atkinson-Shiffrin Memory Model. In *European Conference on Computer Vision*, 640–658.

Cui, S.; Wang, S.; Zhuo, J.; Li, L.; Huang, Q.; and Tian, Q. 2020. Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3941–3950.

Ding, N.; Xu, Y.; Tang, Y.; Xu, C.; Wang, Y.; and Tao, D. 2022. Source-free domain adaptation via distribution estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7212–7222.

Gao, Z.; Huang, K.; Zhang, R.; Liu, D.; and Ma, J. 2023. Towards Better Robustness against Common Corruptions for Unsupervised Domain Adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 18882–18893.

Gretton, A.; Smola, A.; Huang, J.; Schmittfull, M.; Borgwardt, K.; and Schölkopf, B. 2009. *Covariate shift and local learning by distribution matching*, 131–160. Cambridge, MA, USA: MIT Press.

Huang, J.; Guan, D.; Xiao, A.; Lu, S.; and Shao, L. 2022. Category contrast for unsupervised domain adaptation in visual tasks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1203–1214.

Huang, X.; Zhu, C.; and Chen, W. 2023. Semi-supervised Domain Adaptation via Prototype-based Multi-level Learning. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, 884–892.

Hwang, U.; Lee, J.; Shin, J.; and Yoon, S. 2024. SF(DA)^2: Source-free Domain Adaptation Through the Lens of Data Augmentation. In *The Twelfth International Conference on Learning Representations*.

Jing, M.; Li, J.; Lu, K.; Zhu, L.; and Shen, H. T. 2024. Visually Source-Free Domain Adaptation via Adversarial Style Matching. *IEEE Transactions on Image Processing*, 33: 1032–1044.

Karim, N.; Mithun, N. C.; Rajvanshi, A.; Chiu, H.-p.; Samarasekera, S.; and Rahnavard, N. 2023. C-SFDA: A Curriculum Learning Aided Self-Training Framework for Efficient Source Free Domain Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 24120–24131.

Li, J.; Li, G.; Liu, F.; and Yu, Y. 2022. Neighborhood Collective Estimation for Noisy Label Identification and Correction. In *European Conference on Computer Vision*, 128–145.

Li, J.; Li, G.; Shi, Y.; and Yu, Y. 2021a. Cross-Domain Adaptive Clustering for Semi-Supervised Domain Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2505–2514.

Li, J.; Li, G.; and Yu, Y. 2023. Adaptive Betweenness Clustering for Semi-Supervised Domain Adaptation. *IEEE Transactions on Image Processing*, 5580–5594.

Li, K.; Liu, C.; Zhao, H.; Zhang, Y.; and Fu, Y. 2021b. Ecacl: A holistic framework for semi-supervised domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8578–8587.

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 *International Conference on Machine Learning*, 6028–6039.

Litrico, M.; Del Bue, A.; Morerio, P.; et al. 2023. Guiding Pseudo-Labels With Uncertainty Estimation for Source-Free Unsupervised Domain Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7640–7650.

Liu, S.; Liu, K.; Zhu, W.; Shen, Y.; and Fernandez-Granda, C. 2022. Adaptive early-learning correction for segmentation from noisy annotations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2606–2616.

Melas-Kyriazi, L.; and Manrai, A. K. 2021. Pixmatch: Unsupervised domain adaptation via pixelwise consistency training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12435–12445.

Ngo, B. H.; Chae, Y. J.; Kwon, J. E.; Park, J. H.; and Cho, S. I. 2023. Improved Knowledge Transfer for Semi-supervised Domain Adaptation via Trico Training Strategy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 19214–19223.

Peng, X.; Usman, B.; Kaushik, N.; Hoffman, J.; Wang, D.; and Saenko, K. 2017. Visda: The visual domain adaptation challenge. In *arXiv preprint arXiv:1710.06924*.

Saenko, K.; Kulis, B.; Fritz, M.; and Darrell, T. 2010. Adapting visual category models to new domains. In *European Conference on Computer Vision*, 213–226.

Singh, A. 2021. Clda: Contrastive learning for semi-supervised domain adaptation. *Advances in Neural Information Processing Systems*, 34: 5089–5101.

Song, J.; Kim, T. S.; Kim, J.; Nam, G.; Kooi, T.; and Choo, J. 2024. Is User Feedback Always Informative? Retrieval Latent Defending for Semi-supervised Domain Adaptation Without Source Data. In *European Conference on Computer Vision*, 72–92.

Tian, J.; Saddik, A. E.; Xu, X.; Li, D.; Cao, Z.; and Shen, H. T. 2025. Intrinsic Consistency Preservation With Adaptively Reliable Samples for Source-Free Domain Adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, 36(3): 4738–4749.

Venkateswara, H.; Eusebio, J.; Chakraborty, S.; and Panchanathan, S. 2017. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5018–5027.

Wang, S.; Chen, Y.; He, Z.; Yang, X.; Wang, M.; You, Q.; and Zhang, X. 2023a. Disentangled Representation Learning with Causality for Unsupervised Domain Adaptation. In *Proceedings of the 31st ACM International Conference on Multimedia*, 2918–2926.

Wang, S.; Xu, X.; Ma, X.; Jiang, K.; and Wang, Z. 2023b. Informative Classes Matter: Towards Unsupervised Domain Adaptive Nighttime Semantic Segmentation. In *Proceedings of the 31st ACM International Conference on Multimedia*, 163–172.

Wu, Y.; Xing, M.; Zhang, Y.; Xie, Y.; Fan, J.; Shi, Z.; and Qu, Y. 2023. Cross-modal Unsupervised Domain Adaptation for 3D Semantic Segmentation via Bidirectional Fusion-then-Distillation. In *Proceedings of the 31st ACM International Conference on Multimedia*, 490–498.

Xia, H.; and Ding, Z. 2020. Structure preserving generative cross-domain learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4364–4373.

Yan, Z.; Wu, Y.; Li, G.; Qin, Y.; Han, X.; and Cui, S. 2022. Multi-level consistency learning for semi-supervised domain adaptation. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, 1530–1536.

Yang, J.; Shi, S.; Wang, Z.; Li, H.; and Qi, X. 2021a. St3d: Self-training for unsupervised domain adaptation on 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10368–10378.

Yang, L.; Wang, Y.; Gao, M.; Shrivastava, A.; Weinberger, K. Q.; Chao, W.-L.; and Lim, S.-N. 2021b. Deep co-training with task decomposition for semi-supervised domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8906–8916.

Yang, S.; Jui, S.; van de Weijer, J.; et al. 2022. Attracting and dispersing: A simple approach for source-free domain adaptation. *Advances in Neural Information Processing Systems*, 35: 5802–5815.

Yang, S.; van de Weijer, J.; Herranz, L.; Jui, S.; et al. 2021c. Exploiting the intrinsic neighborhood structure for source-free domain adaptation. *Advances in Neural Information Processing Systems*, 34: 29393–29405.

Yi, L.; Xu, G.; Xu, P.; Li, J.; Pu, R.; Ling, C.; McLeod, I.; and Wang, B. 2023. When Source-Free Domain Adaptation Meets Learning with Noisy Labels. In *The Eleventh International Conference on Learning Representations*.

Yoon, J.; Kang, D.; and Cho, M. 2022. Semi-Supervised Domain Adaptation via Sample-to-Sample Self-Distillation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1978–1987.

Yu, Y.-C.; and Lin, H.-T. 2023. Semi-Supervised Domain Adaptation with Source Label Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 24100–24109.

Zhang, J.; Huang, J.; Jiang, X.; and Lu, S. 2023. Black-Box Unsupervised Domain Adaptation with Bi-Directional Atkinson-Shiffrin Memory. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 11771–11782.

Zhang, Z.; Chen, W.; Cheng, H.; Li, Z.; Li, S.; Lin, L.; and Li, G. 2022. Divide and contrast: Source-free domain adaptation via adaptive contrastive learning. *Advances in Neural Information Processing Systems*, 35: 5137–5149.

Zhou, H.-Y.; Chen, X.; Zhang, Y.; Luo, R.; Wang, L.; and Yu, Y. 2022. Generalized radiograph representation learning via cross-supervision between images and free-text radiology reports. *Nature Machine Intelligence*, (1): 32–40.

Zhou, W.; Fan, H.; Luo, T.; and Zhang, L. 2023. Unsupervised Domain Adaptive Detection with Network Stability Analysis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6986–6995.

Zhu, Y.; Zhuang, F.; Wang, J.; Ke, G.; Chen, J.; Bian, J.; Xiong, H.; and He, Q. 2020. Deep subdomain adaptation network for image classification. *IEEE Transactions on Neural Networks and Learning Systems*, (4): 1713–1722.