Cycle Self-Refinement for Multi-Source Domain Adaptation

Chaoyang Zhou[1], **Zengmao Wang***[1,2,3,4,5], **Bo Du**[1,2,3,4,5], **Yong Luo**[1,2,3,4,5]

[1] School of Computer Science, Wuhan University
[2] National Engineering Research Center for Multimedia Software, Wuhan University
[3] Institute of Artificial Intelligence, Wuhan University
[4] Key Laboratory of Multimedia and Network Communication Engineering, Wuhan University
[5] Hubei Luojia Laboratory, China
zhoucy,wangzengmao,dubo,luoyong@whu.edu.cn

Abstract

Multi-source domain adaptation (MSDA) aims to transfer knowledge from multiple source domains to the unlabeled target domain. In this paper, we propose a cycle self-refinement domain adaptation method, which progressively attempts to learn the dominant transferable knowledge in each source domain in a cycle manner. Specifically, several source-specific networks and a domain-ensemble network are adopted in the proposed method. The source-specific networks are adopted to provide the dominant transferable knowledge in each source domain for instance-level ensemble on predictions of the samples in target domain. Then these samples with high-confidence ensemble predictions are adopted to refine the domain-ensemble network. Meanwhile, to guide each source-specific network to learn more dominant transferable knowledge, we force the features of the target domain from the domain-ensemble network and the features of each source domain from the corresponding source-specific network to be aligned with their predictions from the corresponding networks. Thus the adaptation ability of source-specific networks and the domain-ensemble network can be improved progressively. Extensive experiments on Office-31, Office-Home and DomainNet show that the proposed method outperforms the state-of-the-art methods for most tasks.

Introduction

Recently, deep learning has achieved great success in multiple applications, such as image recognition (He et al. 2016), sentiment analysis (Vaswani et al. 2017), and audio processing (Purwins et al. 2019). However, powerful deep-learning techniques typically depend on abundant labeled data. Acquiring such data often involves significant costs, making it inefficient to gather new labeled data for each new scenario. Unsupervised domain adaptation addresses the problem by transferring knowledge from the labeled source domain to the unlabeled target domain. At present, many unsupervised domain adaptation methods (Ganin et al. 2016; Saito et al. 2018; Yang et al. 2023a) that adapt a single source domain to a single target domain have been proposed and achieved great success. Generally, the source domains usually can be collected from multiple environments or sources (Ren et al. 2022). Hence, it is necessary to consider the diversity of

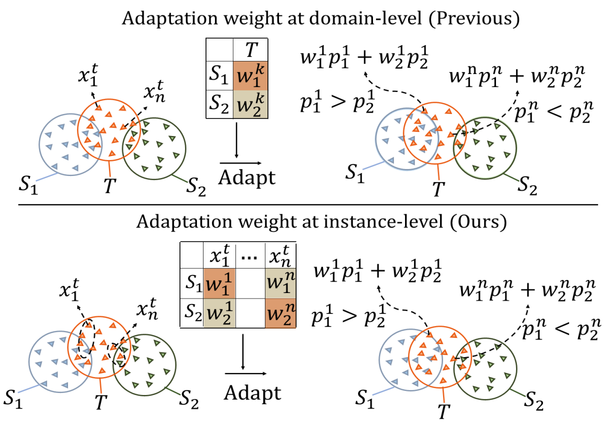


Figure 1: Comparison of previous and the proposed method. S_1, S_2 are two source domains and T is the target domain. Light and dark blocks represent small and large values, respectively. w_i^k and p_i^k are the weight and prediction probability of i^{th} source domain on x_k^t, respectively.

multiple source domains to guarantee the adaptation performance on the target domain, which is a popular learning strategy called Multi-Source Domain Adaptation (MSDA).

To take full advantage of the knowledge from multi-source domains, many strategies have been developed by aggregating the source domains or source models for target adaptation (Li et al. 2021a; Wilson, Doppa, and Cook 2023). For the methods to aggregate the source domains, they usually ensure that the target-relevant source domains are given more importance for feature alignment and classification in a single shared adaptation network (Wen, Greiner, and Schuurmans 2020; Turrisi et al. 2022). Another mainstream paradigm attempts to aggregate multiple source models. These methods adapt each source domain to the target domain separately and then weigh the predictions of multiple source models for inference (Venkat et al. 2020; Shen, Bu, and Wornell 2023). Although previous methods have already achieved competitive results on MSDA, they mainly focus on measuring the importance of each source domain to the samples in the target domain at the domain level, resulting in the partial dominant transferable knowledge of source domains with poorer adaptation ability to the whole target

*Corresponding author.

domain are ignored, limiting the performance of MSDA. As shown in Figure 1, the source domain with poorer adaptation ability can also provide high-confidence predictions for a part of target samples. However, these predictions can not dominate the target prediction due to the domain's small weight for aggregation in previous methods.

To capture the dominant transferable knowledge in each source domain, we propose a *cycle self-refinement* domain adaptation method by aggregating dominant transferable knowledge of source domains at the instance level for each target sample, termed CSR. The method consists of several source-specific networks that separately learn specific knowledge for each source domain. Then an instance-level ensemble strategy is adopted to aggregate the predictions of each source-specific network based on their confidence in the target samples. For a target sample, the ensemble strategy assigns a large weight to the source-specific network that provides confident and consistent predictions on similar samples to the sample. Thus the source-specific network with high-confidence predictions can dominate the ensemble prediction of the target sample. We use these target samples with high-confidence pseudo-labels to train a domain-ensemble network. The domain-ensemble network can not only preserve the high-confidence predictions of the target samples in each source-specific network but also provide more confident pseudo-labels than each source-specific network. Hence, the pseudo-labels provided by the domain-ensemble network are adopted to improve the adaptation ability of source-specific networks. We employ these pseudo-labels of the target domain from the domain-ensemble network and predictions of the source domain from the source-specific network to force the conditional feature alignment between features from their corresponding networks. This creates a cycle mechanism where the domain-ensemble network and the source-specific networks refine each other. The source-specific networks can provide more high-confidence pseudo-labels to improve the domain-ensemble network, while the domain-ensemble network can guide the source-specific networks to adapt more discriminatively. The main contributions can be summarized as:

- We propose a cycle self-refinement method for multiple source domain adaptation. The source-specific networks and the domain-ensemble network are adopted to refine each other by accumulating the adaptation ability of each source domain during the cycle refinement.
- We propose an instance-level ensemble method for target domain pseudo-labeling with multiple source domains. It can discover the dominant transferable knowledge in each source domain and achieve optimal aggregation for each target sample effectively.
- Extensive experiments on Office31, Office-Home and DomainNet show that the proposed method outperforms most of the state-of-the-art methods.

Related Works

Unsupervised Domain Adaptation

Motivated by the theoretical error bound proposed by (Ben-David et al. 2010), many unsupervised domain adaptation methods try to minimize the distribution distance between source and target domain. Some methods minimize the distance measured by data statistics, such as MMD (Long et al. 2018; Zhang et al. 2022) and CORAL (Sun and Saenko 2016). While other methods propose to minimize the distribution distance by adversarial learning (Tzeng et al. 2017; Chen et al. 2022b; Chhabra, Venkateswara, and Li 2023). Some methods also consider fine-grained feature mapping between two domains (Zhou et al. 2023). Although these methods are useful in single-source domain adaptation, they are unsuitable to be applied directly in multi-source domain adaptation due to the gaps among multiple source domains.

Multiple Source Domain Adaptation

With the consideration of the diversity of source domains, most MSDA methods attempt to find a better knowledge aggregation of source domains or models. Some methods (Wen, Greiner, and Schuurmans 2020; Shui et al. 2021; Chen and Marchand 2023) try to exclude less relevant source domains to combine a target-like source domain for easier alignment. But they only consider the weight of source domain at the domain level. While some methods (Zhu, Zhuang, and Wang 2019; Xu et al. 2022) force the prediction consistency between different source models, they overlook the transferability discrepancy of source domains. (Nguyen et al. 2021) adopts a teacher-student framework to distill knowledge of source models to target domain to find the relationship between source models and target samples, but it lacks correction from student to teacher. Recently, the concept of dynamic network has been introduced into this task, which uses a embed machining that changes with samples into the whole model (Li et al. 2021b; Deng et al. 2022). They also embrace dynamic mapping, but the use of extra modules increases the burden of training and inference time.

Self-Training

Self-training is a competitive technique for semi-supervised learning. They always use unlabeled data to regularize the model by training on pseudo-labels. Some methods (Zhang et al. 2021; Sun, Lu, and Ling 2023) generate the pseudo-labels from the current model while other methods (Pham et al. 2021) generate labels from the teacher network. These pseudo-labels are usually used by forcing the strongly augmented samples to have the same pseudo-labels as the weakly augmented ones (Yang et al. 2023b). Recently, a cycle self-training method (Liu, Wang, and Long 2021) and a debiased self-training method (Chen et al. 2022a) use an extra target classifier to train unlabeled data and debias the source model. The proposed method departs from them by circularly refining source-specific networks and a domain-ensemble network with guidance from each other.

The Proposed Method

MSDA adapts the knowledge from multiple labeled source domains to the unlabeled target domain. For convenience, we denote the source domains as $S = [s_1, s_2, ..., s_m]$, where s_i is the i^{th} source domain and m is the number of source domains. While the target domain is denoted as

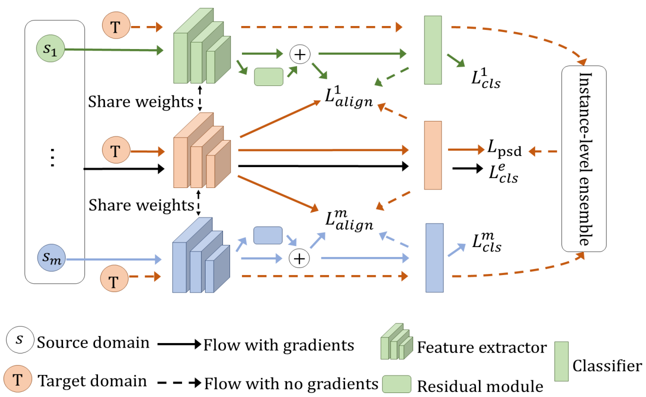


Figure 2: The flowchart of the proposed method. The proposed method uses an instance-level ensemble strategy to aggregate the predictions of source-specific networks for each target sample. These samples with high-confidence pseudo-labels are used to train the domain-ensemble network. The domain-ensemble network is adopted to guide the source-specific networks.

T. The number of categories is denoted as C. In the proposed framework, we employ one backbone network f_c to extract the common features, which can save time and space cost-effectively compared with adopting multiple backbone networks. Then a residual network ϕ_i is used to extract domain-specific features of i^{th} source domain. Thus we use $f_{s_i} = \phi_i \circ f_c$ to represent i^{th} source-domain feature extractor, where $f_1 \circ f_2(x)$ represents $f_1(f_2(x))$. While for target domain to learn the ensemble knowledge from all source domains, we adopt $f_e = f_c$ to exploit the common knowledge of all source domains. Then i^{th} source classifier h_{s_i} and a domain-ensemble classifier h_e are used for classification. Thus $\{f_{s_i}, h_{s_i}\}$ is denoted as i^{th} source-specific network, which is trained with i^{th} source domain. And $\{f_e, h_e\}$ is denoted as domain-ensemble network. After training, we only adopt domain-ensemble network for inference.

Based on the model architecture, we propose an instance-level cycle self-refinement method for multi-source domain adaptation as shown in Figure 2. The proposed method uses the instance-level ensemble strategy to aggregate the predictions of source-specific networks to provide high-confidence pseudo-labels for the training of the domain-ensemble network. While the domain-ensemble network is adopted to guide the refinement of the source-specific networks. In the following, we will introduce the proposed method with instance-level ensemble strategy and the cycle self-refinement learning respectively.

Instance-level Ensemble Strategy

It is still a challenge for MSDA to aggregate the transferable knowledge from multiple source domains for optimal adaptation. For the traditional methods, they measure the adaptation ability of each source domain to the target domain at the domain level and the weight of each source domain for each target sample is the same. In this way, some dominant transferable knowledge in the poor-adaptation source domain may be ignored, since the poor-adaptation source domain usually has a small weight for aggregation. Hence we propose an instance-level ensemble strategy to discover the dominant transferable knowledge from multiple source domains for the prediction of the target sample, which mainly considers the adaptation ability for a target sample not the adaptation ability for the whole target domain.

Inspired by the fact that the good model should have not only high-confidence prediction but also high category prediction diversity on target domain (Cui et al. 2020), we consider the prediction confidence and diversity to estimate the importance of each source-specific network for prediction ensemble. Inspired by cross-entropy, we calculate the prediction confidence of a sample with the prediction of the sample and its similar samples in the target domain. Specifically, for the i^{th} source-specific network, given a target sample x^t, the prediction confidence of x^t with the i^{th} source-specific network can be represented as

$$c^{x^t}_{ic_i} = -H(p^{x^t}_i, \frac{1}{|N^{x^t}|} \sum_{x \in N^{x^t}} p^x_i), \tag{1}$$

where $H(p^1, p^2) = -\sum_{i \in C}(p^1_i logp^2_i + p^2_i logp^1_i)/2$, which measures the similarity and confidence of two predictions simultaneously. $p^x_i = \sigma(h_{s_i} \circ f_e(x))$, where σ is the softmax function. N^{x^t} is the similar sample set for x^t, which is obtained by clustering the target domain into C clusters with the features from the domain-ensemble network based on K-means and then choosing the cluster including x^t as N^{x^t}. Thus the confidence estimation of x^t can be more precise by $c^{x^t}_{ic_i}$, due to that the similar samples of x^t are considered.

While for the prediction diversity, it is mainly used to indicate whether the network predicts diversely. Since the target domain lacks labels, we discover the category structure at the feature level. When the target domain is divided into C clusters, the samples with the same category are assumed to be within the same cluster. Hence, we can average the predictions of the samples in the same cluster to represent the prediction distribution of a category. Then for a sample, we can estimate its prediction diversity by comparing the prediction distribution of all clusters and the cluster that it belongs to. For a sample x^t, its prediction diversity from the i^{th} source-specific network can be represented as:

$$c^{x^t}_{cd_i} = H(\frac{1}{|N^{x^t}|} \sum_{x \in N^{x^t}} p^x_i, \frac{1}{C} \sum_{k \in [C]} \frac{1}{|N^k|} \sum_{x \in N^k} p^x_i), \tag{2}$$

where N^k is the k^{th} cluster when we divide the target domain into C clusters with K-means. We should note that if the source-specific network can predict diversely and have a low preference to the category of x^t, $c^{x^t}_{cd_i}$ is large. The $c^{x^t}_{ic_i}$ and $c^{x^t}_{cd_i}$ are essential to estimate whether the prediction of x^t by the i^{th} source-specific network is reliable.

Moreover, in MSDA, there exist several domains that can provide predictions for the ensemble. Although we attempt to capture the dominant knowledge of each source domain, it is also important to provide consistent predictions with diverse predictions for ensemble learning as indicated in (Opitz and Shavlik 1995). Hence, we use an ensemble diversity regularization to measure the similarity between the prediction of i^{th} source-specific network and the average prediction, which is defined as:

$$r_{d_i}^{x^t} = -H(p_i^{x^t}, p_{avg}^{x^t}), \tag{3}$$

where $p_{avg}^{x^t}$ is the average prediction of x^t by all the source-specifc networks, defined as $p_{avg}^{x^t} = \frac{1}{m}\sum_{i=1}^{m} p_i^{x^t}$. In Eq.(3), the large $r_{d_i}^{x^t}$ indicates that the $p_i^{x^t}$ is useful for ensemble learning. Thus in instance-level ensemble, the importance of i^{th} source-specific network for x^t can be calculated with the summation of prediction confidence, prediction diversity and ensemble diversity regularization, i.e. $c_i^{x^t} = c_{ic_i}^{x^t} + c_{cd_i}^{x^t} + r_{d_i}^{x^t}$. To balance the importance of the source-specific networks for the sample x^t conveniently, the weight is normalized as

$$w^{x^t} = \sigma(< c_1^{x^t}, ..., c_m^{x^t} >). \tag{4}$$

The instant-level ensemble prediction for x^t can be represented as $p_w^{x^t} = \sum_{i=1}^{m} w_i^{x^t} p_i^{x^t}$.

Cycle Self-Refinement Learning

In the proposed method, we use the pseudo-labels provided by the source-specific networks with instance-level ensemble strategy to refine the domain-ensemble network. Meanwhile, the domain-ensemble network is adopted to guide the adaptation of the source-specific networks. Thus the domain-ensemble network and the source-specific networks are improved with each other by a cycle mechanism.

Specifically, to refine source-specific networks, we adopt conditional feature alignment between source features from each source-specific network and target features from domain-ensemble network. In the conditional feature alignment, the pseudo-labels of the target domain are provided by domain-ensemble network while the predictions of the source domain are provided by the corresponding source-specific network. Since the domain-ensemble network benefits from multi-source domains with ensemble learning, it can not only provide precise discriminative distribution in the target domain but also preserve the dominant transferable knowledge in each source domain. Thus the conditional feature alignment between each source domain and the target domain can guide part of the poor adaptation knowledge of one source-specific network tends to be softly consistent with dominant adaptation knowledge from the other source-specific networks. For the conditional feature alignment, similar to Margin Disparity Discrepancy in (Zhang et al. 2019), we align each source domain with the target domain by an auxiliary classifier. The auxiliary classifier is trained with predictions of the source domain and pseudo-labels of the target domain with an adversarial strategy, which is used to guarantee that the samples with the same categories in source domain and target domain can be classified with similar predictions by the source domain classifier. The conditional feature alignment loss between each source domain and target domain can be represented as:

$$\begin{aligned} L_{align}^{i} = \min_{f_{s_i}, f_e} \max_{h'_{s_i}} - \sum_{x^t \in T} CE(1 - \sigma(h'_{s_i} \circ f_e(x^t)), \hat{y}^{x^t}) \\ - \lambda \sum_{x^{s_i} \in s_i} CE(\sigma(h'_{s_i} \circ f_{s_i}(x^{s_i})), \hat{y}^{x^{s_i}}), \end{aligned} \tag{5}$$

where h'_{s_i} is the auxiliary classifier for alignment with i^{th} source domain. $\hat{y}^{x^{s_i}}$ is the prediction of x^{s_i} provided by the i^{th} source-specific network while $\hat{y}^{x^t}$ is the pseudo-label of x^t provided by domain-ensemble network. λ is the trade-off hyper-parameters and set as 3 similar to (Zhang et al. 2019). With the above loss, the predictions of the i^{th} source-specific network and that of the domain-ensemble network can be refined simultaneously. Meanwhile, when the source-specific networks are refined, more high-confident pseudo-labels can be provided with instance-level ensemble for the domain-ensemble network training. Thus the source-specific networks and the domain-ensemble network can be improved progressively with cycle self-refinement.

Total Training Procedure

The proposed method mainly uses the cycle self-refinement manner to train the source-specific networks and the domain-ensemble network. For the source-specific networks, the cross entropy loss is adopted to train each source-specific network. While for the domain-ensemble network, all the source domains are adopted to train. Initially, the classification loss to train the source-specific networks and the domain-ensemble network can be represented as:

$$\begin{aligned} L_{cls} = \sum_i L_{cls}^i + L_{cls}^e = \sum_{i \in [m]} \sum_{x^{s_i} \in s_i} \min_{f_{s_i}, h_{s_i}} CE(p_i^{x^{s_i}}, y^{x^{s_i}}) \\ + \sum_{i \in [m]} \sum_{x^{s_i} \in s_i} \min_{f_e, h_e} CE(p_e^{x^{s_i}}, y^{x^{s_i}}), \end{aligned} \tag{6}$$

where $p_i^{x^{s_i}} = \sigma(h_{s_i} \circ f_{s_i}(x^{s_i}))$ and $p_e^{x^{s_i}} = \sigma(h_e \circ f_e(x^{s_i}))$. y^{s_i} is the true label of x^{s_i} in the i^{th} source domain. Then to exploit the dominant transferable knowledge of each source domain, we employ instance-level ensemble strategy to aggregate predictions of all source-specific networks. We select the target samples with high-confidence pseudo-labels to refine the domain-ensemble network. To further improve the generalization ability of the domain-ensemble network on target domain, motivated by Fixmatch (Sohn et al. 2020), augmentation trick is adopted for consistency regularization. Meanwhile, it is inevitable to provide noisy labels in the pseudo-labels. Hence, for each pseudo-labeled target sample, a weight is adopted to control the classification loss of the pseudo-labels. The domain-ensemble network can be refined with pseudo-labeled target samples by

$$L_{psd} = \min_{f_e, h_e} \sum_{x^t \in T} \gamma^{x^t} (CE(\bar{p}^{x^t}, \hat{y}_w^{x^t}) + JS(p_e^{x^t}, p_a^{x^t})), \tag{7}$$

where $\bar{p}^{x^t} = (p_e^{x^t} + p_a^{x^t})/2$, $p_e^{x^t} = \sigma(h_e \circ f_e(x^t))$ and $p_a^{x^t} = \sigma(h_e \circ f_e(\mathcal{T}(x^t)))$. $\mathcal{T}$ is a strong augmentation operation. $\hat{y}_w^{x^t} = argmax \;\; p_w^{x^t}$. JS is the JS divergence that measures the similarity of two predictions. γ^{x^t} is the weight based on the confidence of $p_w^{x^t}$ and is formulated as:

$$\gamma^{x^t} = \begin{cases} max(p_w^{x^t})^{2-2\tau}, & \text{if} \quad max(p_w^{x^t}) > \tau \\ 0, & \text{others}, \end{cases} \tag{8}$$

where τ is the threshold to select the high-confidence pseudo-labels. When τ is small, most target samples with low confidence may be selected. With the above strategy to weight each pseudo-labeled target sample, the noise taken by the pseudo-labeled target samples with low confidence can be reduced. Meanwhile, the source-specific networks are refined with the conditional feature alignment. Thus the total loss to refine both source-specific networks and the domain-ensemble network is formulated as:

$$L_{all} = L_{cls} + L_{psd} + \beta \sum_{i=1}^{m} L_{align}^{i}, \tag{9}$$

where β is the trade-off hyper-parameter. With the total loss, the source-specific networks and the domain-ensemble network can be trained in an end-to-end manner. During the training, the pseudo-labels of the target samples to refine the source-specific networks and the domain-ensemble network are updated in each iteration.

Theoretical Analysis

We analyze the proposed method theoretically. We derive the target error bound of the domain-ensemble network in Theorem 1 to show how the source-specific networks can refine the domain-ensemble network with the instance-level ensemble. Then combined with the observation of (Zhu et al. 2020), which indicates that the error of one source model on the target domain can be bounded by source classification loss, the accuracy of target pseudo labels for conditional feature alignment, and the domain discrepancy of all categories between two domains, we find that the proposed cycle refinement can bound target domain error of the whole model.

Theorem 1 Suppose there is a data augmentation set $\mathcal{T}$ and $\mathbb{E}_{x^t \in T}(I(argmax \;\; p_e^{x^t} \neq argmax \;\; p_a^{x^t})) \leq \mu$. $I(c)$ is 1 if c is true otherwise 0. Assume T satisfies (q, ϵ)-constant expansion hypothesis, i.e. $q, \epsilon \in (0, 1)$, for any $A \subset T$ and $q < \mathbb{P}(A) < \frac{1}{2}$, when $\mathcal{N}(A) = \{x' || \mathcal{T}(x) - x'| < r, x \in A\}$, we have $\mathbb{P}(\mathcal{N}(A) \backslash A) \geq min\{\mathbb{P}(A), \epsilon\}$. Given $\sum_{j \in [m]} w_j = 1$, we have the classification error of domain-ensemble network as:

$$\begin{aligned} \varepsilon_T(h_e, f_e) &= \mathbb{E}_{x^t \in T}(I(argmax \;\; p_e^{x^t} \neq y^{x^t})) \\ &\leq \sum_{j \in [m]} w_j \varepsilon_T(h_{s_j}, f_e) + l_T(p_e, p_w) \\ &\quad + \frac{\mu}{min\{\epsilon, q\}} + 2q, \end{aligned} \tag{10}$$

The assumption in Theorem 1 is proved in (Wei et al. 2020). In Theorem 1, y^{x^t} is the true label of x^t and $l_T(p_e, p_w) = \mathbb{E}_{x^t \in T}(I(argmax \;\;\; p_e^{x^t} \;\; \neq \;\; argmax \;\;\; p_w^{x^t}))$. The theorem shows that the generalization error of the domain-ensemble network can be bounded by the weighted error of all source-specific networks $\sum_{j \in [m]} w_j \varepsilon_T(h_{s_j}, f_e)$, the distance between predictions of h_e and ensemble predictions $l_T(p_e, p_w)$, and the consistency between different augmented outputs of the samples μ. Given the fixed source-specific networks, the first term can be further reduced with instance-level ensemble, and the last two terms can be reduced with L_{psd}. Suppose domain-level weight can get by averaging the weights of all samples and the covariance between the error and weight of the source-specific network is negative, i.e $w_j = \mathbb{E}_{x \in T}(w_j^x)$ and $Cov_{x \in T}(w_j^x, \varepsilon_x(h_{s_j}, f_e)) < 0$, based on $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) + Cov(X, Y)$, we have

$$\mathbb{E}_{x \in T}(w_j^x \varepsilon_x(h_{s_j}, f_e)) < \sum_{j \in [m]} w_j \varepsilon_T(h_{s_j}, f_e) \tag{11}$$

Hence, the proposed instance-level strategy can provide lower error of ensemble predictions from the source-specific networks than the ensemble predictions at the domain level.

Standards	Methods	$\rightarrow A$	$\rightarrow D$	$\rightarrow W$	Avg
Single Best	DANN	68.2	99.4	96.8	88.1
	MCD	69.7	**100.0**	98.5	89.4
Source Combine	DANN	67.6	99.7	98.1	88.5
	MCD	68.5	99.4	99.3	89.0
Multi-Source	MFSAN	72.7	99.5	98.5	90.2
	SImpAI	70.6	99.2	97.4	89.0
	SSG	71.3	**100.0**	95.5	90.3
	DCA	55.1	99.6	98.9	91.2
	PTMDA	75.4	**100.0**	**99.6**	91.7
	CSR(ours)	**78.6**	**100.0**	**99.6**	**92.7**

Table 1: Classification accuracy(%) on Office-31 dataset.

Experiments

Implementation Details

Three popular benchmark datasets are adopted in the experiments, i.e. Office-31, Office-Home and DomainNet. Office-31 (Saenko et al. 2010) is a classical domain adaptation dataset that has 31 categories and contains threes domains, i.e. Amazon(A), Webcam(W) and Dslr(D); Office-Home (Venkateswara et al. 2017) consists of 65 categories and contains 4 domains, which are Art(Ar), Clipart(Cl), Product(Pr), and Real-World(Rw); DomainNet (Peng et al. 2019) is a larger and more difficult dataset, which contains 0.6 million images with 345 categories in 6 domains, i.e. Clipart(Clp), Infograph(Inf), Painting(Pnt), Quickdraw(Qdr), Real(Rel) and Sketch(Skt). In MSDA experiments, for each dataset, we regard one domain as the target domain while the remaining domains are used as the source domains.

We compare our method with several state-of-the-art methods, such as DANN(Ganin et al. 2016), MCD(Saito et al. 2018), MFSAN(Zhu, Zhuang, and Wang 2019), DCA(Li et al. 2022), PTMDA(Ren et al. 2022), DRT(Li et al. 2021b), WADN (Shui et al. 2021), SImpAI(Venkat

Standards	Methods	$\rightarrow Ar$	$\rightarrow Pr$	$\rightarrow Cl$	$\rightarrow Rw$	Avg
Single Best	DANN	67.9	80.4	55.9	75.8	70.0
	MCD	69.1	79.6	52.2	75.1	69.0
Source Combine	DANN	68.4	79.5	59.1	82.7	72.4
	MCD	67.8	79.2	59.9	80.9	71.9
Multi-Source	MFSAN	72.1	80.3	62.0	81.8	74.1
	DCA	72.1	80.5	63.6	81.4	74.4
	SImpAI	70.8	80.2	56.3	81.5	72.2
	WADN	73.4	86.3	70.2	**87.3**	79.4
	BDT	72.6	85.9	67.4	83.6	77.4
	CSR(ours)	**76.7**	**86.8**	**71.4**	85.5	**80.1**

Table 2: Classification accuracy(%) on Office-Home dataset.

Methods	$\rightarrow Clp$	$\rightarrow Inf$	$\rightarrow Pnt$	$\rightarrow Qdr$	$\rightarrow Rel$	$\rightarrow Skt$	Avg
SImpAI	66.4	26.5	56.6	18.9	68.0	55.5	48.6
SSG	68.7	24.8	55.7	18.4	68.8	56.3	48.8
DRT+ST	71.0	**31.6**	**61.0**	12.3	**71.4**	**60.7**	51.3
MRF-MSDA	63.9	28.7	56.3	16.8	67.1	54.3	47.9
PTMDA	66.0	28.5	58.4	13.0	63.0	54.1	47.2
CSR(ours)	**73.0**	28.1	58.8	**26.0**	71.1	**60.7**	**52.9**

Table 3: Classification accuracy(%) on DomainNet dataset.

et al. 2020), SSG(Yuan et al. 2022), MRF-MSDA(Xu, Wang, and Ni 2022) and BDT(Kundu et al. 2022). Following previous works, we report the experimental results in three aspects: (1)Single Best(SB) shows the highest accuracy among single-source domain adaptation results; (2)Source Combine(SC) shows the accuracy on single-source domain adaptation and the source domain is a combination of all source domains; (3)Multi-Source(MS) shows the performance of multi-source domain adaptation methods.

In the experiments, we use Resnet-50 pre-trained on ImageNet as the backbone for Office31 and Office-Home datasets. The pre-trained Resnet-101 is adopted for DomainNet dataset. When only the source domains are trained (experiments of Figure 3), the Resnet-50 is used for training stability. We utilize the same learning rate and schedule as (Zhu, Zhuang, and Wang 2019). Meanwhile, for trade-off parameters β and filtering threshold τ, we set (0.7, 0.9) for Office31 and Office-home, and (0.7, 0.6) for DomainNet. The RandAugment(Cubuk et al. 2020) is adopted as data augmentation for $\mathcal{T}$. We report the public results for the compared methods. Experiments are done on Nvidia V100. Code is released at https://github.com/zcy866/CSR.

Experimental Results

In the experiments, we run 5 times on each task for the proposed method and report the average results of Office31, Office-Home, and DomainNet in Table 1, Table 2 and Table 3, respectively. We can observe that the performance of single best methods is usually worse than that of the source combined methods. This indicates that different source domains may contain different dominant transferable knowledge. Hence, it is necessary to aggregate transferable knowledge of source domains at instance level to improve the adaptation ability. For the compared multi-source domain adaptation methods, DCA and WADN attempt to use a domain classifier to measure the similarity between each source domain and target domain, and weight multiple source domains for adaptation at domain level. MFSAN and SImpAI adopt predictions to align the source models and then averagely aggregate the multiple source domains for prediction directly. These methods are related to the proposed method. Compared with these state-of-the-art methods, the proposed method has achieved the best performance on all benchmark datasets. This demonstrates that the proposed method can exploit the dominant transferable knowledge from each source domain effectively at instance level and the cycle self-refinement can effectively boost the dominant transferable knowledge learning of all source domains.

Methods	$\rightarrow Ar$	$\rightarrow Pr$	$\rightarrow Cl$	$\rightarrow Rw$	Avg
L_{cls}	68.5	80.3	59.6	81.4	72.4
$L_{cls} + L_{psd}$	76.1	84.6	70.2	84.4	78.8
$L_{cls} + L_{psd} + L_{align}$	76.7	86.8	71.4	85.5	80.1

Table 4: Contribution of each component in the proposed method on the Office-Home dataset.

Methods	$\rightarrow Clp$	$\rightarrow Pnt$	$\rightarrow Qdr$	$\rightarrow Rel$	Avg
$w/o\ c_{ic}$	72.8	58.0	24.6	70.6	56.5
$w/o\ c_{cd}$	72.8	58.4	25.6	70.7	56.9
$w/o\ r_d$	72.9	58.2	25.9	70.9	57.0
p_{avg}	72.4	57.4	24.3	70.2	56.1
p_w	73.0	58.8	26.0	71.1	57.2

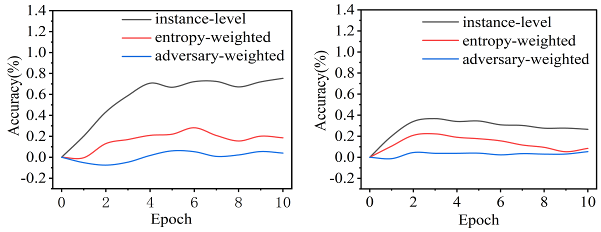
Table 5: Contribution of each component in the instance-level ensemble strategy with four tasks in DomainNet.

Ablation Study

To demonstrate the importance of each component in the proposed method, we show the ablation studies about the variants of the proposed method in Table 4 and Table 5 on Office-Home and DomainNet respectively.

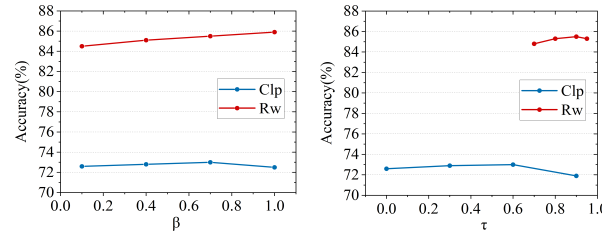
From Table 4, we can observe that when the self-training loss L_{psd} and the conditional feature alignment loss L_{align} are adopted in sequence, the average accuracy increases 6.4% and 1.3% respectively. From the results, we can observe that each component is essential to the proposed cycle mechanism, where the source-specific networks provide highly confident pseudo-labels to improve the domain-ensemble network with L_{psd} while the domain-ensemble network guides the source-specific networks to adapt more discriminative features with L_{align}.

In Table 5, we show the effectiveness of the instance-level ensemble prediction and report the results of its variants. p_{avg} represents the performance with an averaging weighting strategy. p_w represents the performance with instance-level ensemble. From the results, we note that when we replace the instance-level ensemble strategy with the average



(a) $\rightarrow Clp$ (b) $\rightarrow Skt$

Figure 3: The performance with the proposed instance-level ensemble strategy and other domain-level ensemble strategies by subtracting their accuracy and the accuracy of the average-weighted strategy.



(a) (b)

Figure 4: The analysis of hyper-parameters β and τ on Clipart task($\rightarrow Clp$) of in DomainNet dataset and RealWorld task($\rightarrow Rw$) in Office-Home dataset.

weighting ensemble strategy, the average accuracy decreases indicating that the instance-level ensemble successfully exploits knowledge of multiple source-specific networks. Besides, when c_{ic} is removed, the accuracy is reduced, which shows that our confidence measure can capture the dominant predictions of each source-specific network. The accuracy is also reduced when c_{cd} or r_d is removed, indicating the importance of considering the prediction category diversity and ensemble diversity. Hence, the instance-level ensemble strategy is very effective in estimating the confidence of the target sample at the instance level.

To verify the motivation, we show an example of epoch 4 on $\rightarrow Clp$ task in the DomainNet dataset where $\{Qdr$, Inf, Skt, Rel, Pnt $\}$ are used as the source domains. The adaptation accuracies are {31.2%, 31.8%, 47.0%, 49.1%, 37.3% } by using $\{Qdr$, Inf, Skt, Rel, Pnt $\}$ to train source-specific networks. Then we report the proportions of the target samples that are classified correctly by one source-specific network and misclassified by other source-specific networks. The results are {2.7%, 2.0%, 3.1%, 4.3%, 1.5% } when $\{Qdr$, Inf, Skt, Rel, Pnt $\}$ are adopted to train source-specific networks. We can observe that although Qdr has poor adaptation ability, its source-specific network still has 2.7% target samples that can only be classified correctly by the network. Meanwhile, we also report the proportion of the target samples that are correctly classified by both the proposed instance-level ensemble prediction and only one source network. The results are {6.6%, 8.7%, 36.8%, 46.1%, 31.7% } for the source-specific networks with $\{Qdr$, Inf, Skt, Rel, Pnt $\}$ respectively. While the proportions of the target samples classified correctly by both the domain-level ensemble prediction with the weight based on adversarial domain discrepancy (Ganin et al. 2016) and only one source network are {1.9%, 8.4%, 36.0%, 37.0%, 25.8% }. Furthermore, we also compare the proposed instance-level ensemble strategy with the entropy-weighted ensemble and adversary-weighted ensemble at the domain level in Figure 3. In Figure 3, the results show that instance-level ensemble strategy performs better than entropy-weighted ensemble and adversary-weighted ensemble at the domain level. These results indicate that the instance-level ensemble strategy can keep the adaptation ability of each source domain well and is effective in the ensemble of the dominant transferable knowledge in multiple source domains.

Parameters Analysis

In the proposed method, there are two important parameters β and τ. We report the parameters analysis on $\rightarrow Rw$ task of the Office-Home dataset and on $\rightarrow Clp$ task of the DomainNet dataset in Figure 4. We choose β from candidates {0.1, 0.4, 0.7, 1.0} and τ from candidates { 0.7, 0.8, 0.9, 0.95 } for Office-Home and from candidates { 0.0, 0.3, 0.6, 0.9 } for DomainNet. The results show that a large value of β can improve the adaptation performance effectively. It is better to be set as 0.7 in real-world applications. While for τ, it is very different on the Office-Home dataset and DomainNet dataset. The main reason is that the number of categories and samples in DomainNet is much larger than that of Office-Home. When these numbers are small, we can use a large value to choose the high-confidence samples; otherwise, a small value can be adopted.

Conclusion

In this paper, an instance-level cycle self-refinement method for MSDA is proposed by refining the source-specific networks and domain-ensemble network in a cycle manner to aggregate the dominant transferable knowledge in each source domain for adaptation. In the proposed method, an instance-level ensemble strategy is designed to estimate the adaption ability of each source domain for each target sample and is effective in providing high-confidence pseudo-labels to train the domain-ensemble network. Since the domain-ensemble network is trained by the ensemble predictions of the source-specific networks, it can refine the source-specific networks with much more useful knowledge than the knowledge of each source-specific network. With the cycle manner, the domain-ensemble network can be more and more effective for target domain inference. Extensive experiments on the popular benchmark datasets show that the proposed method can outperform most of the state-of-the-art methods.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grants 62271357, 62225113, 62006176, U23A20318, the Natural Science Foundation of Hubei Province under Grants 2023BAB072, the Fundamental Research Funds for the Central Universities under Grants 2042023kf0134.

References

Ben-David, S.; Blitzer, J.; Crammer, K.; Kulesza, A.; Pereira, F.; and Vaughan, J. W. 2010. A theory of learning from different domains. *Machine Learning*, 79: 151–175.

Chen, B.; Jiang, J.; Wang, X.; Wan, P.; Wang, J.; and Long, M. 2022a. Debiased self-training for semi-supervised learning. *Advances in Neural Information Processing Systems*, 35: 32424–32437.

Chen, L.; Chen, H.; Wei, Z.; Jin, X.; Tan, X.; Jin, Y.; and Chen, E. 2022b. Reusing the task-specific classifier as a discriminator: Discriminator-free adversarial domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7181–7190.

Chen, Q.; and Marchand, M. 2023. Algorithm-Dependent Bounds for Representation Learning of Multi-Source Domain Adaptation. In *International Conference on Artificial Intelligence and Statistics*, 10368–10394.

Chhabra, S.; Venkateswara, H.; and Li, B. 2023. Generative Alignment of Posterior Probabilities for Source-free Domain Adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 4125–4134.

Cubuk, E. D.; Zoph, B.; Shlens, J.; and Le, Q. V. 2020. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, 702–703.

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.

Deng, Z.; Zhou, K.; Li, D.; He, J.; Song, Y.-Z.; and Xiang, T. 2022. Dynamic instance domain adaptation. *IEEE Transactions on Image Processing*, 31: 4585–4597.

Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; and Lempitsky, V. 2016. Domain-adversarial training of neural networks. *The Journal of Machine Learning Research*, 17: 2096–2030.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Cision and Pattern Recognition*, 770–778.

Kundu, J. N.; Kulkarni, A. R.; Bhambri, S.; Mehta, D.; Kulkarni, S. A.; Jampani, V.; and Radhakrishnan, V. B. 2022. Balancing discriminability and transferability for source-free domain adaptation. In *International Conference on Machine Learning*, 11710–11728.

Li, K.; Lu, J.; Zuo, H.; and Zhang, G. 2021a. Multi-source contribution learning for domain adaptation. *IEEE Transactions on Neural Networks and Learning Systems*, 33: 5293–5307.

Li, K.; Lu, J.; Zuo, H.; and Zhang, G. 2022. Dynamic classifier alignment for unsupervised multi-source domain adaptation. *IEEE Transactions on Knowledge and Data Engineering*, 35: 4727–4740.

Li, Y.; Yuan, L.; Chen, Y.; Wang, P.; and Vasconcelos, N. 2021b. Dynamic transfer for multi-source domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10998–11007.

Liu, H.; Wang, J.; and Long, M. 2021. Cycle self-training for domain adaptation. *Advances in Neural Information Processing Systems*, 34: 22968–22981.

Long, M.; Cao, Y.; Cao, Z.; Wang, J.; and Jordan, M. I. 2018. Transferable representation learning with deep adaptation networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41: 3071–3085.

Nguyen, V.-A.; Nguyen, T.; Le, T.; Tran, Q. H.; and Phung, D. 2021. Stem: An approach to multi-source domain adaptation with guarantees. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9352–9363.

Opitz, D.; and Shavlik, J. 1995. Generating accurate and diverse members of a neural-network ensemble. *Advances in Neural Information Processing Systems*, 8.

Peng, X.; Bai, Q.; Xia, X.; Huang, Z.; Saenko, K.; and Wang, B. 2019. Moment matching for multi-source domaic:33n adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 1406–1415.

Pham, H.; Dai, Z.; Xie, Q.; and Le, Q. V. 2021. Meta pseudo labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11557–11568.

Purwins, H.; Li, B.; Virtanen, T.; Schlüter, J.; Chang, S.-Y.; and Sainath, T. 2019. Deep learning for audio signal processing. *IEEE Journal of Selected Topics in Signal Processing*, 13: 206–219.

Ren, C.-X.; Liu, Y.-H.; Zhang, X.-W.; and Huang, K.-K. 2022. Multi-source unsupervised domain adaptation via pseudo target domain. *IEEE Transactions on Image Processing*, 31: 2122–2135.

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

Saito, K.; Watanabe, K.; Ushiku, Y.; and Harada, T. 2018. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3723–3732.

Shen, M.; Bu, Y.; and Wornell, G. W. 2023. On Balancing Bias and Variance in Unsupervised Multi-Source-Free Domain Adaptation. In *International Conference on Machine Learning*, 30976–30991.

Shui, C.; Li, Z.; Li, J.; Gagné, C.; Ling, C. X.; and Wang, B. 2021. Aggregating from multiple target-shifted sources. In *International Conference on Machine Learning*, 9638–9648.

Sohn, K.; Berthelot, D.; Carlini, N.; Zhang, Z.; Zhang, H.; Raffel, C. A.; Cubuk, E. D.; Kurakin, A.; and Li, C.-L. 2020. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in Neural Information Processing Systems*, 33: 596–608.

Sun, B.; and Saenko, K. 2016. Deep coral: Correlation alignment for deep domain adaptation. In *Computer Vision–ECCV 2016 Workshops*, 443–450.

Sun, T.; Lu, C.; and Ling, H. 2023. Domain adaptation with adversarial training on penultimate activations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 9935–9943.

Turrisi, R.; Flamary, R.; Rakotomamonjy, A.; and Pontil, M. 2022. Multi-source domain adaptation via weighted joint distributions optimal transport. In *Uncertainty in Artificial Intelligence*, 1970–1980.

Tzeng, E.; Hoffman, J.; Saenko, K.; and Darrell, T. 2017. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7167–7176.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Venkat, N.; Kundu, J. N.; Singh, D.; Revanur, A.; et al. 2020. Your classifier can secretly suffice multi-source domain adaptation. *Advances in Neural Information Processing Systems*, 33: 4647–4659.

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

Wei, C.; Shen, K.; Chen, Y.; and Ma, T. 2020. Theoretical Analysis of Self-Training with Deep Networks on Unlabeled Data. In *International Conference on Learning Representations*.

Wen, J.; Greiner, R.; and Schuurmans, D. 2020. Domain aggregation networks for multi-source domain adaptation. In *International Conference on Machine Learning*, 10214–10224.

Wilson, G.; Doppa, J. R.; and Cook, D. J. 2023. Calda: Improving multi-source time series domain adaptation with contrastive adversarial learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–14.

Xu, M.; Wang, H.; and Ni, B. 2022. Graphical modeling for multi-source domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–1.

Xu, Y.; Kan, M.; Shan, S.; and Chen, X. 2022. Mutual learning of joint and separate domain alignments for multi-source domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1890–1899.

Yang, J.; Liu, J.; Xu, N.; and Huang, J. 2023a. Tvt: Transferable vision transformer for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 520–530.

Yang, L.; Qi, L.; Feng, L.; Zhang, W.; and Shi, Y. 2023b. Revisiting weak-to-strong consistency in semi-supervised semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7236–7246.

Yuan, J.; Hou, F.; Du, Y.; Shi, Z.; Geng, X.; Fan, J.; and Rui, Y. 2022. Self-supervised graph neural network for multi-source domain adaptation. In *Proceedings of the 30th ACM International Conference on Multimedia*, 3907–3916.

Zhang, B.; Wang, Y.; Hou, W.; Wu, H.; Wang, J.; Okumura, M.; and Shinozaki, T. 2021. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. *Advances in Neural Information Processing Systems*, 34: 18408–18419.

Zhang, Y.; Liu, T.; Long, M.; and Jordan, M. 2019. Bridging theory and algorithm for domain adaptation. In *International Conference on Machine Learning*, 7404–7413.

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, Q.; Gu, Q.; Pang, J.; Lu, X.; and Ma, L. 2023. Self-adversarial disentangling for specific domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45: 8954–8968.

Zhu, Y.; Zhuang, F.; and Wang, D. 2019. Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 5989–5996.

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*, 32: 1713–1722.