Discriminability and Transferability Estimation: A Bayesian Source Importance Estimation Approach for Multi-Source-Free Domain Adaptation

Zhongyi Han[1], Zhiyan Zhang[1], Fan Wang[1], Rundong He[1], Wan Su[1], Xiaoming Xi[2], Yilong Yin[1]*

[1]School of Software, Shandong University
[2]School of Computer Science and Technology, Shandong Jianzhu University
hanzhongyicn@gmail.com, ylyin@sdu.edu.cn

Abstract

Source free domain adaptation (SFDA) transfers a single-source model to the unlabeled target domain without accessing the source data. With the intelligence development of various fields, a zoo of source models is more commonly available, arising in a new setting called *multi-source-free domain adaptation* (MSFDA). We find that the critical inborn challenge of MSFDA is how to estimate the importance (contribution) of each source model. In this paper, we shed new Bayesian light on the fact that the posterior probability of source importance connects to discriminability and transferability. We propose Discriminability And Transferability Estimation (DATE), a universal solution for source importance estimation. Specifically, a proxy discriminability perception module equips with habitat uncertainty and density to evaluate each sample's surrounding environment. A source-similarity transferability perception module quantifies the data distribution similarity and encourages the transferability to be reasonably distributed with a domain diversity loss. Extensive experiments show that DATE can precisely and objectively estimate the source importance and outperform prior arts by non-trivial margins. Moreover, experiments demonstrate that DATE can take the most popular SFDA networks as backbones and make them become advanced MSFDA solutions.

Introduction

Learning machines, especially deep neural networks (DNNs), show proficiency in multiple arrays of real-world tasks under stationary environments where we draw training (source) and test (target) examples from an identical distribution (LeCun, Bengio, and Hinton 2015; Redmon et al. 2016; Huang et al. 2017). However, many studies in theory and practice have demonstrated that learning machines fail to generalize even if the source and target distributions slightly differ (Valiant 1984; Acuna et al. 2021). To guarantee generalizability under non-stationary environments, unsupervised domain adaptation (UDA) has gained momentum in the past decade with prominent theoretical advances and effective algorithms (Han et al. 2020; Courty et al. 2017). Thanks to the free access to source data, the UDA algorithms for learning domain-invariant representations yield state-of-the-art performance on many visual tasks (Tzeng et al. 2017).

*Corresponding auhtor.

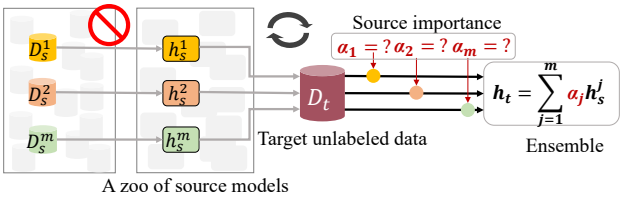


Figure 1: Problem setup of MSFDA. The key is to estimate the importance (contribution degree) of each source model.

Source data are often distributed on edge devices and carry private information, *e.g.*, those on medical instruments or from corporate financial data (Liang, Hu, and Feng 2020). *Source-free domain adaptation* (SFDA) relaxes the requirement on the source data and aims to transfer a previously trained source model instead of the source data to the unlabeled target domain. A series of seminal studies have achieved significant success on effective algorithms for SFDA variants, such as *white-box* SFDA (Kim, Cho, and Hong 2020; Liang, Hu, and Feng 2020; Wang et al. 2022a), *black-box* SFDA (Nelakurthi, Maciejewski, and He 2018; Sahoo, Shanmugam, and Guttag 2020), *class-mismatch* SFDA (Kundu et al. 2020), and *active* SFDA (Wang et al. 2022b).

In real-world applications, a zoo of well-trained source models is more easily obtainable under the umbrella of multiple decentralized sources and privacy protection (Feng et al. 2021). Accordingly, a new SFDA variant termed *multi-source-free domain adaptation* (MSFDA) is emerged, adapting multi-source models to the unlabeled target domain, as illustrated in Fig. 1. MSFDA enables transferring adequate source knowledge and obtaining more increased performance than SFDA. MSFDA has also more practical significance in real-world applications, such as object recognition (Feng et al. 2021), semantic segmentation (He et al. 2021), person re-identification (Wu et al. 2019; Ding, Duan, and Li 2022), *etc*. MSFDA is under-explored as only two general studies have made a solid step (Ahmed et al. 2021; Dong et al. 2021).

A new and inborn challenge of MSFDA is *how to accurately estimate the importance of each source model in a zoo (briefly described as source importance)?* The reason is that some source models are helpful while others are helpless or negatively influence the target learning. Pioneering MSFDA studies show that selecting the best source model and treat-

ing each source domain as an equal contributor are naive solutions (Ahmed et al. 2021; Dong et al. 2021). Further, conventional multi-source domain adaptation (MSDA) methods cannot be applied to estimate the source importance without accessing the source domain data (Guo, Shah, and Barzilay 2018; Yang et al. 2020a; Li et al. 2018). While pioneering MSFDA studies have recognized the necessity of source importance estimation, their estimated results of source importance extremely mismatch the actual contributions of source models, resulting in unsatisfactory performance.

In this paper, we shed new Bayesian light on source importance and find that the prior probability connects to the discriminability of source models, and the likelihood connects to the transferability of source models on the target domain. We propose the Discriminability And Transferability Estimation (DATE) framework to quantify the posterior probability of source importance objectively and effectively. Accordingly, DATE has two novel targeted modules: **(1)** the proxy discriminability perception module can objectively estimate the discriminability of source models by two newly-designed metrics: habitat uncertainty and habitat density. The key insight is that, rather than estimating the uncertainty of each sample point itself, it is better to consider the proxy environmental uncertainty of the sample habitat in the feature space. **(2)** The source-similarity transferability perception module can effectively estimate the transferability of source models by learning the data distribution similarity across domains. It leverages a multi-layer perception to distinguish the target feature representations extracted by which source model. Moreover, a domain diversity loss is designed to encourage transferability to be reasonably distributed.

We summarize our contributions as follows:

- We investigate a limited-explored problem, multi-source-free domain adaptation, and propose a novel and universal framework called DATE.
- We propose two new metrics called habitat uncertainty and habitat density to support the proxy discriminability perception module to evaluate the discriminability of source models from the perspective of sample habitat.
- We propose the source-similarity transferability perception module by quantifying the similarity degree of target data to the unavailable source data and balancing the transferability and diversity of source models.
- We carry out extensive experiments on four benchmark datasets, demonstrating that our proposal achieves remarkable improvements compared with previous methods.

Methodology

In this section, we introduce the necessary notations. Then, we give an in-depth analysis of the key ingredients of a good source model and the insights of achieving a comprehensive source importance estimation. Further, we present the DATE framework. The framework is illustrated in Fig. 2.

Learning Setup

Definition 1 (**Multi-Source-Free Domain Adaptation**). Given a zoo of m well-trained source models $\mathcal{H}_s = \{h_s^j\}_{j=1}^m$, the j^{th} model h_s^j: $\mathcal{X} \to \mathcal{Y}$ is a classification model trained using the j^{th} source domain dataset $D_s^j \sim p^j$. $\mathcal{Y} \in \{1, ..., K\}$ where K represents the class number. The distributions of m source dataset $\{p^j\}_{j=1}^m$ are different. Note that the source datasets $\{D_s^j\}_{j=1}^m$ cannot be accessed during adaptation. Without losing generality, a general deep source model h_s^j is a two-part network: a feature extractor f: $\mathcal{X} \to \mathcal{Z}$, and a classifier g: $\mathcal{Z} \to \mathcal{Y}$. $\mathcal{Z}$ denotes the feature space that plays a key role in deep unsupervised domain adaptation. The goal of MSFDA is to adapt and aggregate the m source models $h_t = \sum_{j=1}^m \alpha_j h_s^j$ on the target data $D_t = \{x_i\}_{i=1}^{n_t} \sim q$ with satisfactory performance. α_j represents the weight (importance) of the j-th model and $\alpha_j \in \boldsymbol{\alpha} = \{\alpha_j\}_{j=1}^m$.

In-depth Analysis

What are the key ingredients of a good source model? We denote by $u : \mathcal{X} \to \mathcal{Y}$ the target labeling function and $L : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ the loss function defined over pairs of labels. We denote $\epsilon_p(h, u) = \mathbb{E}_{x\sim p} L(h(x), u(x))$ the expected source risk and $\epsilon_q(h, u) = \mathbb{E}_{x\sim q} L(h(x), u(x))$ the expected target risk. Ben-David et al. (2007) and Blitzer et al. (2007) proposed a theoretical upper bound on the expected target risk $\epsilon_q(h, u_q)$. For any hypothesis $h \in \mathcal{H}$, the bound of target expected risk $\epsilon_q(h, u_q)$ is given by

$$\begin{aligned} &\epsilon_q(h, u_q) \le \epsilon_p(h, u_p) + disc_L(p, q) + \lambda(p, q)\,, \text{where} \\ &disc_L(p, q) = \sup_{h,h' \in \mathcal{H}} \left|\epsilon_p(h, h') - \epsilon_q(h, h')\right| \text{and} \\ &\lambda(p, q) = \min_{h\in\mathcal{H}} \epsilon_p(h, u_p) + \epsilon_q(h, u_q)\,. \end{aligned} \tag{1}$$

Here, $disc_L(p, q)$ denotes the discrepancy distance term that represents the measure of distribution discrepancy between the source distribution and target distribution. $\lambda(p, q)$ represents the joint optimal error, *i.e.*, the error of an ideal hypothesis on both source and target domains.

A discriminability and transferability perspective. The discriminability of a source model refers to the ability to distinguish different categories on the source and target domains (Chen et al. 2019). The transferability of a source model refers to the ability to generalize to other different domains. The joint optimal error $\lambda(p, q)$ corresponds to the discriminability of a source model on the source and target domains because the smaller the optimal error, the higher the classification accuracy, and thus the better the discriminability of the model, *i.e.*, the discriminability metric $\gamma_D = 1 - \lambda(p, q)$. The discrepancy distance $disc_L(p, q)$ corresponds to the transferability of a source model to the target domain because the smaller the discrepancy distance, the more transferability to the target domain, *i.e.*, the transferability metric $\gamma_T = 1 - disc_L(p, q)$. Further, the studies in the literature (Chen et al. 2019; Kundu et al. 2022) have proved that exclusively improving the transferability leads to a drop in discriminability and vice versa, *i.e.*, they are at odds with each other. Thus, the increase of discriminability γ_D would decrease the transferability γ_T. Based on the above analysis, we draw the following insights.

Insight 1. A good source model should have high discriminability and transferability with a reasonable tradeoff.

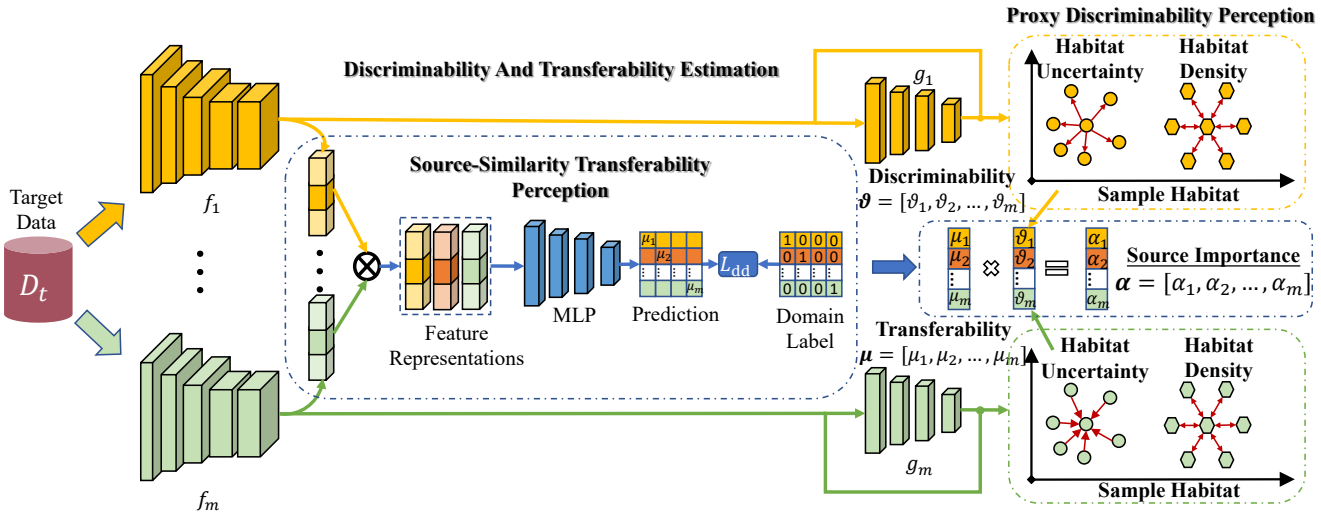


Figure 2: An overview of DATE, mainly including a proxy discriminability perception strategy to quantify the discriminability of source models without target labels, and a source-similarity transferability perception strategy to quantify the transferability.

Remark 1. Since transferability and discriminability have a tradeoff, transferability and discriminability are the two most essential components of a sound source model w.r.t. MSFDA, and one cannot be missing. Next, there is still a question to be analyzed: *What are the connections between transferability or discriminability and the importance of the source model?* We analyze the source importance from the Bayesian perspective to give a comprehensive answer.

A Bayesian perspective. The core of the MSFDA framework is to assign a set of importance $\boldsymbol{\alpha} = \{\alpha_j\}_{j=1}^m$ to each source model, and $\sum_{j=1}^m \alpha_j = 1, 0 \leq \alpha_j \leq 1$. From the Bayesian perspective, given a zoo of source models $\mathcal{H}_s$ and a set of unlabeled target data D_t, the objective of MSFDA is to maximize the posterior probability of target labels:

$$\log P(\mathcal{Y}|\mathcal{H}_s, D_t) = \log \int P(\mathcal{Y}|\boldsymbol{\alpha}, H_s, D_t) P(\boldsymbol{\alpha}|\mathcal{H}_s, D_t) d\boldsymbol{\alpha} \,. \tag{2}$$

$P(\mathcal{Y}|\boldsymbol{\alpha}, H_s, D_t)$ indicates that $\boldsymbol{\alpha}$ impacts the learning of target labels, which is intuitive. The second term $P(\boldsymbol{\alpha}|\mathcal{H}_s, D_t)$ plays a key role in the source importance estimation. Thus, we further dissect the second term as follows.

$$P(\boldsymbol{\alpha}|\mathcal{H}_s, D_t) = \frac{P(\boldsymbol{\alpha}, \mathcal{H}_s, D_t)}{P(\mathcal{H}_s, D_t)} = \frac{P(\boldsymbol{\alpha}|D_t) P(\mathcal{H}_s|\boldsymbol{\alpha}, D_t)}{P(\mathcal{H}_s)} \,. \tag{3}$$

Since $P(H_s, D_t) = P(H_s)P(D_t)$, the denominator is $P(H_s)$, which is uniform, *i.e.*, the probability of occurrence is equal for each source model. Thus the posterior probability of source importance $P(\boldsymbol{\alpha}|\mathcal{H}_s, D_t)$ is approximated to

$$P(\boldsymbol{\alpha}|\mathcal{H}_s, D_t) \propto P(\boldsymbol{\alpha}|D_t) P(\mathcal{H}_s|\boldsymbol{\alpha}, D_t) \,. \tag{4}$$

$P(\boldsymbol{\alpha}|D_t)$ is the prior source importance conditioned on the target data. It represents the latent prior knowledge entailed in the target data D_t w.r.t. source models $\mathcal{H}_s$. Since we cannot access the source data, we can view the unlabeled target data as proxy information to evaluate the discriminability of source models and further inflect the corresponding source importance. $P(\mathcal{H}_s|\boldsymbol{\alpha}, D_t)$ is the likelihood of source models conditioned on the source importance and target data. Assume the feature representations $\mathcal{Z}_t$ of target data are extracted by source models. $P(\mathcal{H}_s|\boldsymbol{\alpha}, D_t)$ indicates the similarity between the source and target data because if the target data has the highest similarity to the j-th source data, the probability of $\mathcal{Z}_t$ being extracted by the j-th source model is the largest. Therefore, $P(\mathcal{H}_s|\boldsymbol{\alpha}, D_t)$ can reflect the transferability. To summarize, the posterior probability of source importance is approximated to

$$P(\boldsymbol{\alpha}|\mathcal{H}_s, D_t) \propto \text{Discriminability} \times \text{Transferability} \,. \tag{5}$$

Accordingly, we would better simultaneously estimate the discriminability and transferability of source models on the target domain. However, a key question remains: *How do we estimate the discriminability and transferability objectively?* The discriminability metric γ_D and transferability metric γ_T are impossible to estimate directly. First, without the target labels, we cannot directly compute the performance on the target domain and quantify the discriminability metric γ_D of source models. To solve this challenge, we propose the proxy discriminability perception module to objectively evaluate the discriminability based on the unlabeled target data. Second, without the source data, we cannot directly compute the discrepancy distance and measure the transferability metric γ_T. To solve this challenge, we propose the novel source-similarity transferability perception module.

Proxy Discriminability Perception

Motivation. Without the target label, entropy is widely used to measure uncertainty from the perspective of information theory. For any source model h_s^j, the self-entropy of a target instance x_i is defined by

$$E(x_i) = -\sum_{k=1}^{K} [\eta \circ g(f(x_i))]_k \log [\eta \circ g(f(x_i))]_k \,, \tag{6}$$

and the self-entropy of target data D_t is

$$E(D_t) = \frac{1}{n}\sum_{i=1}^{n} E(x_i)\,, \tag{7}$$

where η denotes the softmax function and symbol $\circ$ denotes function composition. The lower the entropy, the more unambiguous cluster assignments, and the lower the discriminability (Hu et al. 2017). While self-entropy is intuitive as a generic metric, it poses some limitations. First, self-entropy exhibits low discriminability for highly uncertain and extremely sharp predictions (Fu et al. 2020). Second, when the source model has an over-confidence or under-confidence problem, self-entropy does not truly reflect the actual model discriminability. To avoid these limitations, we turn to estimate the uncertainty of sample habitat because the sample habitat can accurately reflect the actual uncertainty of each target instance and further can reflect the discriminability.

Definition 2 (**Sample Habitat**)**.** Sample habitat is a place where an instance lives in nature. Based on the feature space $\mathcal{Z} = \{z_1, z_2, .., z_{n_t}\}$ of target data and the cosine similarity measurement, we define the samples that are close to x_i in the feature space as the neighbors of x_i: $N(x_i) = \{x_1^i, x_2^i, \cdots, x_q^i\}$, where q denotes the number of the close neighbors. Accordingly, three key ingredients (dimensions) build the sample habitat: instance habitat $\Xi_{\mathcal{X}}$, uncertainty habitat Ξ_E, and distance habitat $\Xi_{\mathcal{D}}$:

$$\begin{aligned}
\Xi_{\mathcal{X}} &= \{x_i, x_1^i, x_2^i, \cdots, x_q^i\},\\
\Xi_E &= \{E_i, E_1^i, E_2^i, \cdots, E_q^i\}, \text{ and}\\
\Xi_D &= \{d_{i,i}, d_{i,1}, d_{i,2}, \cdots, d_{i,q}\},
\end{aligned} \tag{8}$$

where E_q^i represents the entropy of the neighbor x_q^i, and $d_{i,q}$ represents the cosine distance from it q-th neighbor.

Sample habitat can reasonably evaluate the target instance uncertainty in high-dimensional feature space by the inherent structure of the target features and provides a new perspective to evaluate the sample uncertainty from the sample environment. Intuitively, if the sample habitat has low uncertainty, *i.e.*, the sample has the same pseudo-label as its most neighbors, the sample uncertainty is low. In contrast, the sample in the label-chaotic habitat has large uncertainty. Based on this intuitive insight, we design the habitat uncertainty.

Definition 3 (**Habitat Uncertainty**)**.** Based on the sample habitat, for any source model h_s^j, we define the habitat uncertainty $HU(x_i)$ of the i-th instance as

$$HU(x_i) = E_i + \frac{1}{q}\sum_{a=1}^{q} E_a^i, \text{ s.t., } E_a^i \in \Xi_E\,, \tag{9}$$

and the habitat uncertainty of all target data D_t is

$$HU(D_t) = \frac{1}{n}\sum_{i=1}^{n} HU(x_i)\,. \tag{10}$$

In essence, the habitat uncertainty of target data quantifies the entropy of reliable samples repeatedly identified as neighbors of other samples, appearing in multiple sample habitats. Since the uncertainties of the reliable samples are objective, the entire habitat uncertainty can reflect the true discriminability of every source model. However, habitat uncertainty can be susceptible to interference from outliers if a fixed range of habitats contains few samples. To avoid this problem, we propose habitat density as a strong complement.

Definition 4 (**Habitat Density**)**.** Based on the sample habitat, for any source model h_s^j, we define the habitat density $HD(x_i)$ of the i-th instance as

$$HD(x_i) = d_{i,i} + \frac{1}{q}\sum_{a=1}^{q} d_{i,a}, \text{ s.t., } d_{i,a} \in \Xi_D\,, \tag{11}$$

and the habitat density of all target data D_t is

$$HD(D_t) = \frac{1}{n}\sum_{i=1}^{n} HD(x_i)\,. \tag{12}$$

Habitat density describes how close a sample is to its neighbors and measures the compactness of feature space. From an angle, habitat density reveals the intra-class distance, *i.e.*, if the neighbors with the same labels have the closest distances, the decision boundary of each class will be easily found. Therefore, the higher the habitat density of the entire target data, the higher the source model's discriminability.

Since the larger the discriminability, the higher the habitat density and the lower the habitat uncertainty, the proxy discriminability $\boldsymbol{\vartheta}(h_s^j)$ of a source model h_s^j is defined by

$$\boldsymbol{\vartheta}(h_s^j) = \frac{HD(D_t)}{HU(D_t)}\,. \tag{13}$$

Source-Similarity Transferability Perception

Motivation. We have two key principles in the design: **(1)** Each sample of the target domain should have a different similarity from the source data since each sample holds a disparity discrepancy with the source data (Wang et al. 2022b). **(2)** The transferability perception module should be biased toward the source model with high prior importance of $P(\boldsymbol{\alpha}|D_t)$ since $\boldsymbol{\alpha}$ is the conditional information of $P(\mathcal{H}_s|\boldsymbol{\alpha}, D_t)$.

In light of this, the feature representation z_i^j of a target sample x_i extracted by the j-th source model h_s^j can be considered as the representative information of h_s^j, which is then employed as network input to quantify the transferability. We concatenate the feature representations of a batch sample together to obtain $\boldsymbol{\mathcal{Z}}_b^j = [z_1^j, z_2^j, \cdots, z_b^j] \in \mathbb{R}^{b\times c}$. $\boldsymbol{\mathcal{Z}}_b^j$ is then forwarded into a Multi-Layer Perceptron (MLP) network $\boldsymbol{\Gamma}$ to quantify the transferability $\boldsymbol{\mu}(h_s^j)$ of the source model:

$$\boldsymbol{\mu}(h_s^j) = \frac{1}{b}\sum_{i=1}^{b}[\eta \circ \boldsymbol{\Gamma}(z_i^j)]_j\,, \tag{14}$$

and the transferability $\boldsymbol{\mu}$ of a zoo of source models is $\boldsymbol{\mu} = [\boldsymbol{\mu}(h_s^1), \boldsymbol{\mu}(h_s^2), \cdots, \boldsymbol{\mu}(h_s^m)]$, such that $\sum_{j=1}^{m}\boldsymbol{\mu}(h_s^j) = 1$. $[\eta \circ \boldsymbol{\Gamma}(z_i^j)]_j$ represents the probability of the feature representation z_i^j extracted by the j-th source model. If the sample x_i is most similar to the source data, its feature representation

Method	Source Data	Office-31				Office-Caltech				
		R→W	R→D	R→A	Avg.	R→W	R→D	R→A	R→C	Avg.
Source only	✔	97.1	92.0	51.6	80.2	93.5	94.2	90.6	87.5	91.5
MDAN	✔	99.2	95.4	55.2	83.3	99.4	98.7	93.5	91.6	95.8
DCTN	✔	99.6	96.9	54.9	83.8	99.3	99.4	94.1	91.3	96.0
M3SDA	✔	99.4	96.2	55.4	83.7	99.5	99.2	94.5	92.2	96.4
MDDA	✔	99.2	97.1	56.2	84.2	99.3	99.6	95.3	92.3	96.6
LtC-MSDA	✔	99.6	97.2	56.9	84.6	99.4	**99.7**	93.7	95.1	97.0
Source model	✗	95.4	97.5	60.2	84.4	98.0	99.5	96.3	92.1	96.5
BAIT	✗	98.5	98.8	71.1	89.5	98.0	97.5	**97.5**	95.7	97.2
PrDA	✗	93.8	96.7	73.2	87.9	97.6	97.1	97.3	94.6	96.7
SHOT	✗	94.9	97.8	75.0	89.3	99.6	96.8	95.7	95.8	97.0
MA	✗	96.1	97.3	75.2	89.5	99.8	97.2	95.7	95.6	97.1
NRC	✗	95.9	97.9	72.4	88.7	99.3	97.5	95.9	94.9	96.9
DECISION	✗	97.9	98.6	75.3	90.6	99.3	96.8	95.6	95.4	96.8
CAiDA	✗	97.1	99.7	72.7	89.8	99.7	98.1	95.2	95.6	97.1
DATE (NRC)	✗	98.1	**99.8**	73.6	90.5	99.3	98.1	95.9	94.9	97.1
DATE (SHOT)	✗	**99.8**	99.6	**76.4**	**91.9**	**99.8**	98.1	95.6	**95.7**	**97.3**

Table 1: Results on Office-31 and Office-Caltech (ResNet-50). R is the rest domains.

z_i^j extracted by the j-th source model has a larger probability to have a large $\boldsymbol{\mu}(h_s^j)$ than other source models.

While $\mathbf{\Gamma}$ can be optimized by integrating its outputs into the optimization objective as shown in Eq. (18), transferability collapse often appears in some difficult tasks, *i.e.*, the transferability of a source model tends to be one and the other to be zero. Thus, a domain diversity loss L_{dd} is designed for encouraging the outputs to be reasonably distributed and enlarging the diversity of feature representations. To be specific, the domain label of a target sample x_i via the j-th source model feature extractor is formulated as $v_i^j = j$. The collection of domain labels $[\{v_i^1\}_{i=1}^b, \cdots, \{v_i^j\}_{i=1}^b, \cdots, \{v_i^m\}_{i=1}^b] \in \mathbb{R}^{m\times b}$ represents a kind of ground matrix associated with all source models, encoding the unique domain characterization. Given a set of prior source importance (*i.e.*, the proxy discriminability $\{\boldsymbol{\vartheta}(h_s^j)\}_{j=1}^m$), L_{dd} aims to make $\mathbf{\Gamma}$ distinguish its inputs from which source model:

$$L_{\text{dd}} = \frac{1}{m\times b}\sum_{j=1}^{m}\sum_{i=1}^{b}\boldsymbol{\vartheta}(h_s^j)v_i^j\log(\eta\circ\mathbf{\Gamma}(z_i^j))\,, \tag{15}$$

where $\boldsymbol{\vartheta}(h_s^j)$ plays a weighting role to make the module biased toward the source model with high prior importance.

Guided by Eq. (5), the source importance of j-th source model is estimated by combining the proxy discriminability and the source-similarity transferability:

$$\alpha_j = \boldsymbol{\vartheta}(h_s^j)\times\boldsymbol{\mu}(h_s^j)\,. \tag{16}$$

We normalize $\boldsymbol{\alpha} = \{\alpha_j\}_{j=1}^m$ of a zoo of source models, such that $\sum_{j=1}^m \alpha_j = 1, 0 \le \alpha_j \le 1$.

Universal Decision and Optimization

Decision process. Similar to previous MSFDA studies (Ahmed et al. 2021; Dong et al. 2021), we take a convex combination of source models with source importance $\boldsymbol{\alpha}$ to obtain the target predictor:

$$\hat{y}_i = \arg\max\sum_{j=1}^{m}\alpha_j\eta\circ g(f(x_i))\,. \tag{17}$$

Optimization goal. Given m source models $\{h_s^j\}_{j=1}^m$, and target data $\mathcal{D}_t$, we optimize over the parameters $\{\phi_s^j\}_{j=1}^m$ of source models and the parameters $\phi_{\mathbf{\Gamma}}$ of source-similarity transferability perception with source importance. With a hyper-parameter β, the final objective is

$$\begin{aligned}\underset{\{\phi_s^j\}_{j=1}^m,\phi_{\mathbf{\Gamma}},\{\alpha_j\}_{j=1}^m}{\text{minimize}}\quad & L_{\text{backbone}} + \beta L_{\text{dd}} \\ \text{subject to}\quad & 0\le\alpha_j\le 1, \forall j\in\{1,2,\cdots,m\}\,, \\ & \sum_{j=1}^{m}\alpha_j = 1\,.\end{aligned} \tag{18}$$

In practice, L_{backbone} represents the whole loss function of backbone methods that are integrated into our framework. The source importance $\boldsymbol{\alpha}$ is also embedded into the optimization process of backbones and plays weighting roles in the feature combination and decision combination processes via the above convex combination (Eq.(17)), such as pseudo-labeling, information maximization (Liang, Hu, and Feng 2020), neighborhood clustering (Wang et al. 2022b), *etc*.

Method	Source Data	R→ AR	R→ CL	R→ PR	R→ RW	Avg.
Source only	✔	53.4	51.8	71.3	67.8	61.1
MDAN	✔	65.4	62.2	77.6	77.3	70.6
DCTN	✔	66.4	63.8	78.3	78.7	71.8
M3SDA	✔	67.2	63.5	79.1	79.4	72.3
MDDA	✔	66.7	62.3	79.5	79.6	71.0
LtC-MSDA	✔	67.4	**64.1**	79.2	80.1	72.7
Source model only	✗	50.9	50.1	78.8	76.3	64.0
BAIT	✗	71.1	59.6	79.4	77.2	71.8
PrDA	✗	69.3	57.5	79.1	76.8	70.7
SHOT-best	✗	72.1	57.2	83.4	81.3	73.5
SHOT	✗	72.2	59.3	82.8	82.9	74.3
MA	✗	72.5	57.4	82.3	81.7	73.5
NRC	✗	72.7	58.1	82.3	82.1	73.8
DECISION	✗	73.3	58.7	82.9	**84.0**	74.7
CAiDA	✗	70.3	55.0	83.0	80.7	72.2
DATE (NRC)	✗	73.3	58.3	82.3	82.5	74.1
DATE (SHOT)	✗	**75.2**	60.9	**85.2**	**84.0**	**76.3**

Table 2: Results on Office-Home. R is the rest domains.

Experiment

We evaluate DATE on four datasets against state-of-the-art methods. The code is at https://github.com/zhyhan/DATE.

Setup

Datasets To verify the feasibility of the proposed learning method in MSFDA, we thoroughly evaluate the performance of DATE on four benchmark datasets. **Office-31** (Saenko et al. 2010) is a standard DA dataset consisting of three distinct domains: **A**mazon, **W**ebcam, and **D**SLR. It has 4,652 images with 31 unbalanced classes. Extended from the Office-31 dataset, **Office-Caltech** (Gong et al. 2012) includes four domains: **A**mazon, **W**ebcam, **D**SLR, and **C**altech256, in which each domain has 10 categories. **Office-Home** (Venkateswara et al. 2017) is a more challenging DA dataset consisting of 15,599 images with 65 unbalanced classes and four more distinct domains: **Ar**tistic images, **Cl**ip Art images, **Pr**oduct images, and **R**eal-**w**orld images. **Digits-Five** contains five practical domains: MNIST (**MN**), SVHN (**SV**), USPS (**US**), MNIST-M (**MM**), and Synthetic Digits (**SY**).

Baselines We compare our designed Discriminability And Transferability Estimation (DATE) algorithm against multiple state-of-the-art methods: **(1) MSDA**: MDAN (Zhao et al. 2018), DCTN (Wang et al. 2019), M^{3}SDA (Peng et al. 2019), MDDA (Zhao et al. 2020) and LtC-MSDA (Wang et al. 2020). **(2) SFDA**: BAIT (Yang et al. 2020b), PrDA (Kim et al. 2021), SHOT (Liang, Hu, and Feng 2020), NRC (Yang et al. 2021), and MA (Li et al. 2020). **(3) MSFDA**: DECISION (Ahmed et al. 2021), CAiDA (Dong et al. 2021). Note that the MSDA baselines are trained with accessing source data. We extend the SFDA baselines by averaging the predictions of all adapted source models following the pioneering works (Dong et al. 2021). Furthermore, we also build two supervised learning baselines: Source Only (He et al. 2016) combines source data as a training set and views the target data as a test set, and Source Model Only takes an average over the predictions of all source models. Finally, we take SHOT as the backbone of DATE (*i.e.*, DATE-SHOT). Note that the results of DECISION and CAiDA are implemented by ourselves according to the public source code strictly.

Results

Table 1 reports the results on Office-31 and Office-Caltech. Our algorithm significantly outperforms all compared methods. Table 2 and Table 3 report the effects on Office-Home and Digits-Five, where we make a remarkable performance boost. After taking SHOT as the backbone, our algorithm outperforms SHOT by a large margin, showing the necessity of source importance. These results also confirm that our method possesses both simplicity and performance strength.

We find several impressive results: **(1)** As shown in Table 2, SHOT-best (*i.e.*, selecting the best source model) underperforms the SHOT ensemble method (*i.e.*, taking a uniform average), verifying that selecting the best source model is a naive solution. **(2)** As shown in Table 1, the SFDA ensemble methods underperform the MSFDA methods, confirming that treating each source domain as an equal contributor is not a practical solution. **(3)** As shown in all the tasks, our algorithm achieves state-of-the-art results compared to the other MSFDA methods, demonstrating that the estimation of discriminability and transferability of source models can enhance the performance. **(4)** In most tasks, DATE outperforms MSDA methods that require the source data when adaptation.

Method	Source Data	R → MM	R → MT	R → UP	R → SV	R → SY	Avg.
Source only	✔	63.4	90.5	88.7	63.5	82.4	77.7
MDAN	✔	69.5	98.0	92.4	69.2	87.4	83.3
DCTN	✔	70.5	96.2	92.8	77.6	86.8	84.8
M3SDA	✔	72.8	98.4	96.1	81.3	89.6	87.7
MDDA	✔	78.6	98.8	93.9	79.3	89.7	88.1
LtC-MSDA	✔	85.6	**99.0**	**98.3**	**83.2**	93.0	**91.8**
Source model only	✗	25.2	90.0	93.3	42.8	77.8	65.8
BAIT	✗	87.6	96.2	96.7	60.6	90.5	86.3
PrDA	✗	86.2	95.4	95.8	57.4	84.8	83.9
SHOT	✗	90.4	98.9	97.7	58.3	83.9	85.8
MA	✗	**90.8**	98.4	98.0	59.1	84.5	86.2
NRC	✗	74.9	97.6	94.6	73.5	73.5	82.8
DECISION	✗	85.6	98.6	98.0	69.3	95.2	89.4
CAiDA	✗	83.2	98.2	97.8	68.7	94.3	88.4
DATE (NRC)	✗	73.0	98.1	96.0	85.6	89.7	88.5
DATE (SHOT)	✗	86.5	98.6	98.2	73.8	**97.5**	90.9

Table 3: Results on Digits-Five (LeNet).

Method	R → AR	R → CL	R → PR	R → RW	Avg.
DATE (SHOT)	**75.2**	**60.9**	**85.2**	**84.0**	**76.3**
w/o domain diversity loss	74.2	59.8	84.5	82.4	75.2
w/o discriminability	73.7	59.0	84.5	83.6	75.2
w/o habitat uncertainty	74.3	59.9	83.9	83.7	75.4
w/o habitat density	74.7	60.3	82.9	89.9	75.5
w/o transferability	73.7	59.8	82.7	83.5	74.9

Table 4: Ablation study on Office-Home (ResNet-50).

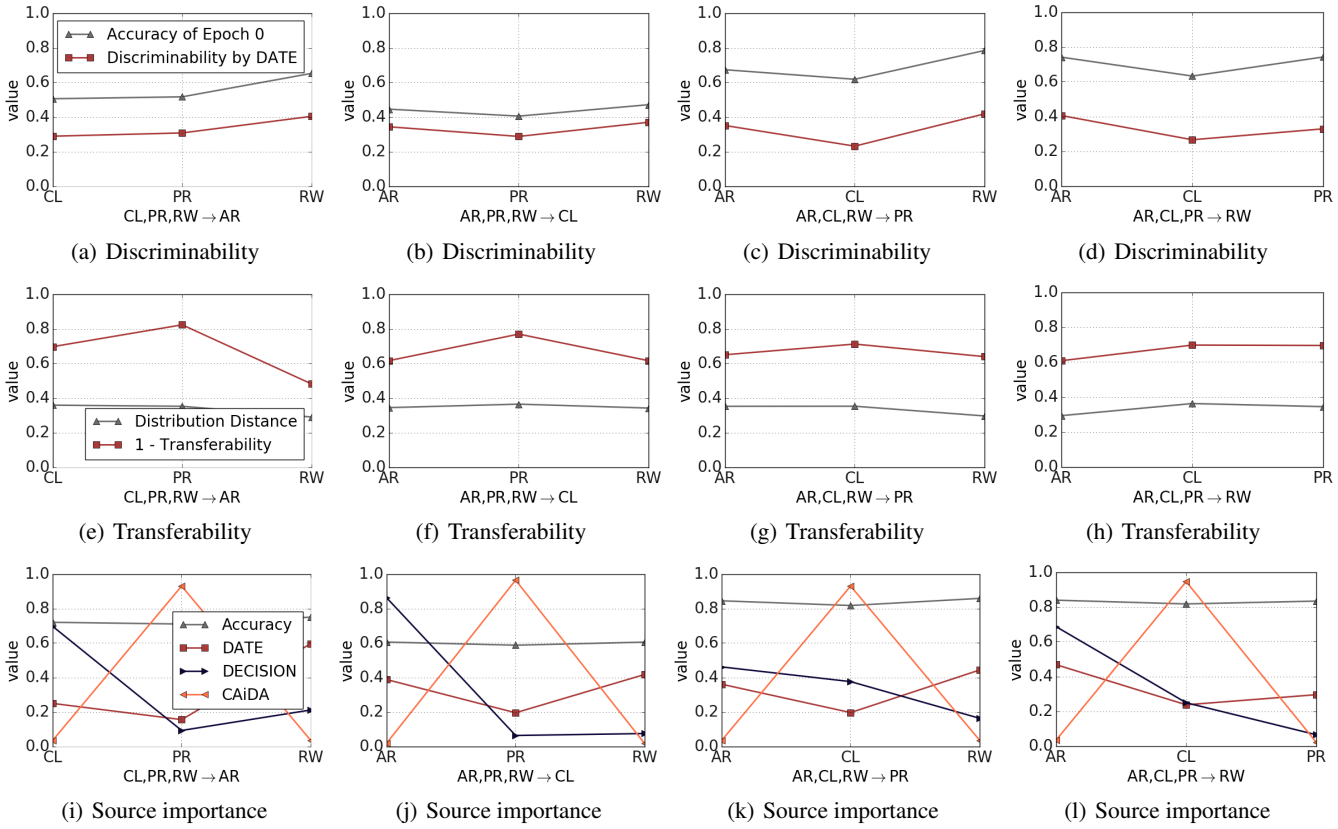
Analyses

Universality analysis. As stated above, we incorporate SHOT into the DATE framework on the four datasets to demonstrate the effectiveness of DATE. To further confirm the universality of DATE, we also conduct experiments on all four datasets by incorporating another type of SFDA method NRC into the DATE framework (**DATE-NRC**). As depicted in Tables 1, 2, 3, DATE-NRC remarkably outperforms NRC on all the MSFDA tasks, showing its universality.

Ablation analysis. We dissect the efficacy of the proposed method by evaluating the variants of DATE on Office-Home as shown in Table 4. DATE remarkably exceeds these variants, proving the necessity of the corresponding modules. **(1)** DATE w/o discriminability is the variant without using proxy discriminability perception. DATE w/o habitat uncertainty or density is the variant without the specific metric. DATE outperforms each variant, indicating the contribution of the novel habitat uncertainty and habitat density to enable accurate discriminability evaluation. **(2)** DATE w/o transferability is the variant without using the source-similarity transferability perception. **(3)** DATE w/o domain diversity loss is the variant without the loss to resolve the transferability collapse and enlarge the diversity of source models.

Discriminability analysis. Fig. 3(a-d) shows the perfect consistent relationship between the estimated discriminability and the actual accuracy of source models without fine-tuning on the target domain (epoch is 0). This result gives strong empirical evidence that the proxy discriminability perception can accurately estimate the discriminability of source models. Fig. 4 displays the t-SNE embeddings (Donahue et al. 2014) of the learned features by DATE and source models on the two tasks of Office-Home. While the features learned by source models are mixed up, the features of each class learned by DATE are more compact, which verifies that our algorithm can learn more discriminative representations.

Transferability analysis. The distribution distance value is a gold standard to represent the actual transferability of source models. Fig. 3(e-h) shows the consistent relationship between the distribution distance and the transferability estimated by the source-similarity transferability perception module. The distribution distance is measured by the proxy $\mathcal{A}$-distance based on the feature representations extracted

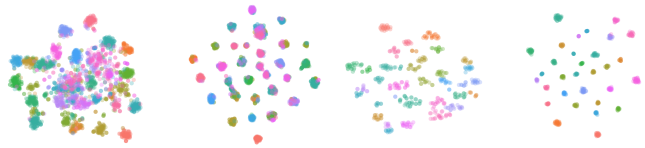


(a) Discriminability (b) Discriminability (c) Discriminability (d) Discriminability

(e) Transferability (f) Transferability (g) Transferability (h) Transferability

(i) Source importance (j) Source importance (k) Source importance (l) Source importance

Figure 3: Our estimated discriminability, transferability, and source importance strictly match the true contributions of source models. The horizontal axis represents the tasks of the Office-Home, while the vertical axis represents the metric value.



(a) Source model(W) (b) DATE (W) (c) Source model (D) (d) DATE (D)

Figure 4: The t-SNE visualization of target data features.

by ImageNet pre-trained model, which supports objective assessment (Han, Sun, and Yin 2022). The results confirm that source-similarity transferability perception can measure the similarity of target data to the source data end-to-end.

Source importance analysis. Fig. 3(i-l) shows the source importance estimated by DATE and the other MSFDA methods, DECISION, and CAiDA, on the Office-Home dataset. We can see that the source importance estimated by our method strictly matches the actual contribution of the source model on the four tasks. The source importance estimated by the CAiDA is opposite to the source model's contribution, while the DECISION cannot accurately estimate source importance. These intuitive results verify the necessity and advantage of simultaneously evaluating the discriminability and transferability of source models.

Conclusion

This paper emphasizes the new and critical problem of source importance estimation to achieve robust multi-source-free domain adaptation. We shed new light on source importance estimation that the prior probability (w.r.t. discriminability) and the likelihood (w.r.t. transferability) of source domains contribute to the source model importance estimation comprehensively and objectively. Therefore, we propose the Discriminability And Transferability Estimation framework with two novel tailored modules. Extensive experiments on four usual datasets have demonstrated that our method could achieve more accurate estimation in various real-world applications. Our approach is simple and orthogonal to other methods. In future work, we believe that our method opens up new possibilities in the topics of multi-source-free domain adaptation.

Acknowledgments

The authors wish to thank the anonymous area chair and reviewers for their helpful comments and suggestions. This work is supported by the National Natural Science Foundation of China (62176139) and the Major Basic Research Project of the Natural Science Foundation of Shandong Province (ZR2021ZD15).

References

Acuna, D.; Zhang, G.; Law, M. T.; and Fidler, S. 2021. f-Domain Adversarial Learning: Theory and Algorithms. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, 66–75. PMLR.

Ahmed, S. M.; Raychaudhuri, D. S.; Paul, S.; Oymak, S.; and Roy-Chowdhury, A. K. 2021. Unsupervised Multi-Source Domain Adaptation Without Access to Source Data. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, 10103–10112. Computer Vision Foundation / IEEE.

Ben-David, S.; Blitzer, J.; Crammer, K.; and Pereira, F. 2007. Analysis of representations for domain adaptation. In *Advances in neural information processing systems*, 137–144.

Blitzer, J.; Crammer, K.; Kulesza, A.; Pereira, F.; and Wortman, J. 2007. Learning Bounds for Domain Adaptation. In *Advances in Neural Information Processing Systems, Vancouver, British Columbia, Canada, December 3-6, 2007*, 129–136. Curran Associates, Inc.

Chen, X.; Wang, S.; Long, M.; and Wang, J. 2019. Transferability vs. Discriminability: Batch Spectral Penalization for Adversarial Domain Adaptation. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97, 1081–1090. PMLR.

Courty, N.; Flamary, R.; Habrard, A.; and Rakotomamonjy, A. 2017. Joint distribution optimal transportation for domain adaptation. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 3733–3742. Red Hook, NY, USA: Curran Associates Inc. ISBN 978-1-5108-6096-4.

Ding, Y.; Duan, Z.; and Li, S. 2022. Source-free unsupervised multi-source domain adaptation via proxy task for person re-identification. *Vis. Comput.*, 38(6): 1871–1882.

Donahue, J.; Jia, Y.; Vinyals, O.; Hoffman, J.; Zhang, N.; Tzeng, E.; and Darrell, T. 2014. DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition. In *Proceedings of the 31th International Conference on Machine Learning*, 647–655.

Dong, J.; Fang, Z.; Liu, A.; Sun, G.; and Liu, T. 2021. Confident Anchor-Induced Multi-Source Free Domain Adaptation. In *Advances in Neural Information Processing Systems, NeurIPS 2021, December 6-14, 2021, virtual*, 2848–2860.

Feng, H.; You, Z.; Chen, M.; Zhang, T.; Zhu, M.; Wu, F.; Wu, C.; and Chen, W. 2021. KD3A: Unsupervised Multi-Source Decentralized Domain Adaptation via Knowledge Distillation. In *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, 3274–3283. PMLR.

Fu, B.; Cao, Z.; Long, M.; and Wang, J. 2020. Learning to Detect Open Classes for Universal Domain Adaptation. In *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XV*, volume 12360 of *Lecture Notes in Computer Science*, 567–583. Springer.

Gong, B.; Shi, Y.; Sha, F.; and Grauman, K. 2012. Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012*, 2066–2073. IEEE Computer Society.

Guo, J.; Shah, D. J.; and Barzilay, R. 2018. Multi-Source Domain Adaptation with Mixture of Experts. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, 4694–4703. Association for Computational Linguistics.

Han, Z.; Gui, X.; Cui, C.; and Yin, Y. 2020. Towards Accurate and Robust Domain Adaptation under Noisy Environments. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, 2269–2276. ijcai.org.

Han, Z.; Sun, H.; and Yin, Y. 2022. Learning Transferable Parameters for Unsupervised Domain Adaptation. *IEEE Transactions on Image Processing*, 1–1.

He, J.; Jia, X.; Chen, S.; and Liu, J. 2021. Multi-Source Domain Adaptation With Collaborative Learning for Semantic Segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, 11008–11017. Computer Vision Foundation / IEEE.

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

Hu, W.; Miyato, T.; Tokui, S.; Matsumoto, E.; and Sugiyama, M. 2017. Learning Discrete Representations via Information Maximizing Self-Augmented Training. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, 1558–1567. PMLR.

Huang, G.; Liu, Z.; van der Maaten, L.; and Weinberger, K. Q. 2017. Densely Connected Convolutional Networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, 2261–2269. IEEE Computer Society.

Kim, Y.; Cho, D.; Han, K.; Panda, P.; and Hong, S. 2021. Domain Adaptation Without Source Data. *IEEE Trans. Artif. Intell.*, 2(6): 508–518.

Kim, Y.; Cho, D.; and Hong, S. 2020. Towards Privacy-Preserving Domain Adaptation. *IEEE Signal Process. Lett.*, 27: 1675–1679.

Kundu, J. N.; Kulkarni, A.; Bhambri, S.; Mehta, D.; Kulkarni, S.; Jampani, V.; and Babu, R. V. 2022. Balancing Discriminability and Transferability for Source-Free Domain Adaptation. *CoRR*, abs/2206.08009.

Kundu, J. N.; Venkat, N.; V., R. M.; and Babu, R. V. 2020. Universal Source-Free Domain Adaptation. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, 4543–4552. IEEE.

LeCun, Y.; Bengio, Y.; and Hinton, G. 2015. Deep learning. *nature*, 521(7553): 436–444.

Li, R.; Jiao, Q.; Cao, W.; Wong, H.-S.; and Wu, S. 2020. Model adaptation: Unsupervised domain adaptation without source data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9641–9650.

Li, Y.; Murias, M.; Dawson, G.; and Carlson, D. E. 2018. Extracting Relationships by Multi-Domain Matching. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, 6799–6810.

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 *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, 6028–6039. PMLR.

Nelakurthi, A. R.; Maciejewski, R.; and He, J. 2018. Source Free Domain Adaptation Using an Off-the-Shelf Classifier. In *IEEE International Conference on Big Data, Big Data 2018, Seattle, WA, USA, December 10-13, 2018*, 140–145. IEEE.

Peng, X.; Bai, Q.; Xia, X.; Huang, Z.; Saenko, K.; and Wang, B. 2019. Moment Matching for Multi-Source Domain Adaptation. In *2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019*, 1406–1415. IEEE.

Redmon, J.; Divvala, S. K.; Girshick, R. B.; and Farhadi, A. 2016. You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, 779–788. IEEE Computer Society.

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. Springer.

Sahoo, R.; Shanmugam, D.; and Guttag, J. V. 2020. Unsupervised Domain Adaptation in the Absence of Source Data. *CoRR*, abs/2007.10233.

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

Valiant, L. G. 1984. A theory of the learnable. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, 436–445. ACM.

Venkateswara, H.; Eusebio, J.; Chakraborty, S.; and Panchanathan, S. 2017. Deep hashing network for unsupervised domain adaptation. In *2018 IEEE Conference on Computer Vision and Pattern Recognition*, 5018–5027.

Wang, F.; Han, Z.; Gong, Y.; and Yin, Y. 2022a. Exploring Domain-Invariant Parameters for Source Free Domain Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7151–7160.

Wang, F.; Han, Z.; Zhang, Z.; and Yin, Y. 2022b. Active Source Free Domain Adaptation. *CoRR*, abs/2205.10711.

Wang, H.; Xu, M.; Ni, B.; and Zhang, W. 2020. Learning to Combine: Knowledge Aggregation for Multi-source Domain Adaptation. In *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part VIII*, volume 12353 of *Lecture Notes in Computer Science*, 727–744. Springer.

Wang, L.; Xu, S.; Wang, X.; and Zhu, Q. 2019. Eavesdrop the Composition Proportion of Training Labels in Federated Learning. *CoRR*, abs/1910.06044.

Wu, A.; Zheng, W.; Guo, X.; and Lai, J. 2019. Distilled Person Re-Identification: Towards a More Scalable System. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, 1187–1196. Computer Vision Foundation / IEEE.

Yang, L.; Balaji, Y.; Lim, S.; and Shrivastava, A. 2020a. Curriculum Manager for Source Selection in Multi-source Domain Adaptation. In *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XIV*, volume 12359 of *Lecture Notes in Computer Science*, 608–624. Springer.

Yang, S.; Wang, Y.; van de Weijer, J.; Herranz, L.; and Jui, S. 2020b. Unsupervised Domain Adaptation without Source Data by Casting a BAIT. *CoRR*, abs/2010.12427.

Yang, S.; Wang, Y.; van de Weijer, J.; Herranz, L.; and Jui, S. 2021. Exploiting the Intrinsic Neighborhood Structure for Source-free Domain Adaptation. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, 29393–29405.

Zhao, H.; Zhang, S.; Wu, G.; Moura, J. M. F.; Costeira, J. P.; and Gordon, G. J. 2018. Adversarial Multiple Source Domain Adaptation. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, 8568–8579.

Zhao, S.; Wang, G.; Zhang, S.; Gu, Y.; Li, Y.; Song, Z.; Xu, P.; Hu, R.; Chai, H.; and Keutzer, K. 2020. Multi-Source Distilling Domain Adaptation. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, New York, NY, USA, February 7-12, 2020*, 12975–12983. AAAI Press.