More is Better: Deep Domain Adaptation with Multiple Sources

Sicheng Zhao[1] , **Hui Chen**[1] , **Hu Huang**[3*] , **Pengfei Xu**[4] , **Guiguang Ding**[1,2]

[1]BNRist, Tsinghua University, China
[2]School of Software, Tsinghua University, China
[3]Peking University Shenzhen Graduate School, China
[4]Didi Chuxing, China

{schzhao, huichen, dinggg}@tsinghua.edu.cn, h.huang@pku.edu.cn, xupengfeipf@didiglobal.com

Abstract

In many practical applications, it is often difficult and expensive to obtain large-scale labeled data to train state-of-the-art deep neural networks. Therefore, transferring the learned knowledge from a separate, labeled source domain to an unlabeled or sparsely labeled target domain becomes an appealing alternative. However, direct transfer often results in significant performance decay due to *domain shift*. Domain adaptation (DA) aims to address this problem by aligning the distributions between the source and target domains. Multi-source domain adaptation (MDA) is a powerful and practical extension in which the labeled data may be collected from multiple sources with different distributions. In this survey, we first define various MDA strategies. Then we systematically summarize and compare modern MDA methods in the deep learning era from different perspectives, followed by commonly used datasets and a brief benchmark. Finally, we discuss future research directions for MDA that are worth investigating.

1 Background and Motivation

The availability of large-scale labeled training data has enabled deep neural networks (DNNs) to achieve remarkable success in many learning tasks, ranging from computer vision to natural language processing [Wilson and Cook, 2020; Zhao *et al.*, 2022]. However, in many practical applications, obtaining sufficient labeled data is often expensive, time-consuming, and even impossible [Zhao *et al.*, 2023]. For example, only experts can provide reliable annotations for fine-grained categories; annotating one Cityscapes image on the pixel level takes about 90 minutes [Zhao *et al.*, 2022].

One potential solution is to transfer the model trained on a separate, labeled source domain to the desired unlabeled or sparsely labeled target domain. But as Figure 1 illustrates, the **direct transfer of models across domains might lead to poor performance**, which is often termed negative transfer [Zhang *et al.*, 2023]. Figure 1(a) shows that even for the simple task of digit recognition, training a LeNet-5 model on

---

*Corresponding Author.

	Accuracy on target
Train on source	**52.3**
Train on target	**96.0**

Source: MNIST Target: MNIST-M

(a) Digit recognition based on LeNet-5

	mIoU on target
Train on source	**21.7**
Train on target	**62.6**

Source: GTA Target: Cityscapges

Train on source   Train on target   Ground truth

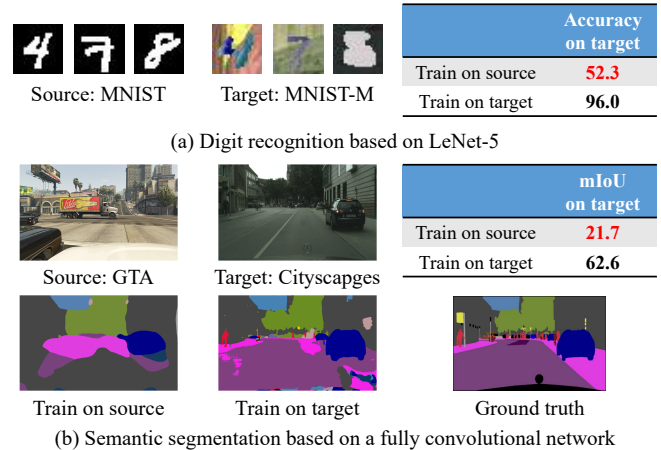(b) Semantic segmentation based on a fully convolutional network

Figure 1: Examples of *domain shift* in the single-source scenario. The models trained on the labeled source domain do not perform well when directly transferring to the target domain.

the MNIST source and directly transferring to the MNIST-M target leads to a classification accuracy decrease from 96.0% to 52.3%. Figure 1(b) shows a more realistic example of training a semantic segmentation model on a synthetic source dataset GTA and conducting pixel-wise segmentation on a real target dataset Cityscapes. The poor performance of direct transfer across domains stems from *domain shift* [Zhao *et al.*, 2022]: the joint probability distributions of observed data and labels are different in the two domains.

The phenomenon of domain shift motivates the research on domain adaptation (DA), which aims to learn a model from a labeled source domain that can generalize well to a different but related target domain. Existing DA methods mainly focus on the single-source scenario. In the deep learning era, recent single-source DA (SDA) methods usually employ a conjoined architecture with two branches to respectively represent the models for the source and target domains. One branch aims to learn a task model based on the labeled source data using corresponding task losses, such as cross-entropy loss for classification. The other branch aims to deal with the domain shift by aligning the target and source domains. Based on the alignment strategies, deep SDA methods can be classified into different categories. The two most popular
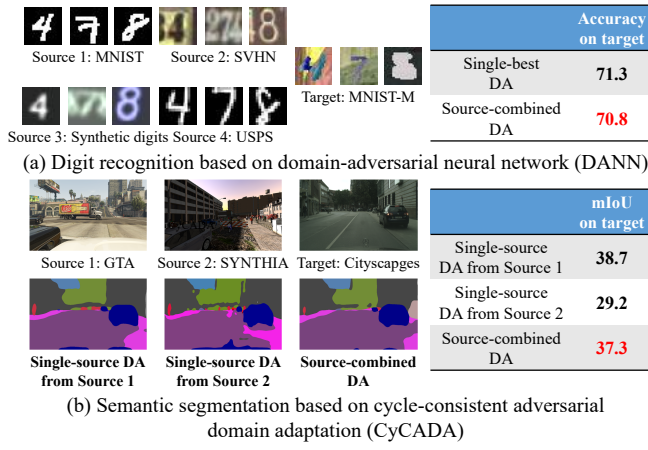
	Accuracy on target
Single-best DA	71.3
Source-combined DA	70.8

(a) Digit recognition based on domain-adversarial neural network (DANN)

	mIoU on target
Single-source DA from Source 1	38.7
Single-source DA from Source 2	29.2
Source-combined DA	37.3

(b) Semantic segmentation based on cycle-consistent adversarial domain adaptation (CyCADA)

Figure 2: Examples of *domain shift* in the multi-source scenario. Simply combining multiple source domains into one source and directly performing SDA does not guarantee better performance compared to just using the best individual source.

categories are 1) learning domain-invariant feature representations through discrepancy-based methods, adversarial discriminative methods, or self-supervision-based methods and 2) generating intermediate domains through generative models, such as Generative Adversarial Network (GAN). Other categories include ensemble methods, target discriminative methods, and so on. Please refer to [Wilson and Cook, 2020; Zhao *et al.*, 2022] for more details.

In practice, **the labeled data may be collected from multiple sources with different distributions** [Sun *et al.*, 2015]. In such cases, the aforementioned SDA methods could be trivially applied by combining multiple sources into a single one: an approach we refer to as *source-combined DA*. However, source-combined DA might result in poorer performance than simply using one of the sources. As illustrated in Figure 2, the accuracy on the best single-source digit recognition adaptation using DANN [Ganin *et al.*, 2016] is 71.3%, while the source-combined accuracy drops to 70.8%. Because the domain shift not only exists between each source and the target but also exists among different sources, the source-combined data may interfere with each other during the learning process [Zhao *et al.*, 2021a]. Therefore, multi-source domain adaptation (MDA) is required to effectively leverage all the available data from different sources.

The early MDA methods mainly focus on shallow models [Sun *et al.*, 2015], either learning a latent feature space for different domains or combining pre-learned source classifiers. Recently, the emphasis on MDA has shifted to deep learning architectures. In this paper, we systematically survey recent progress on deep learning-based MDA, summarize and compare the correlations and differences among these approaches, and discuss potential future directions.

2 Problem Definition

In the typical MDA setting, there are multiple source domains $S_1, S_2, \cdots, S_M$ (M is the number of sources) and one target domain T. Suppose the observed data and corresponding la-

bels in the ith source S_i drawn from distribution $p_i(\mathbf{x}, \mathbf{y})$ are $\mathbf{X}_i = \{\mathbf{x}_i^j\}_{j=1}^{N_i}$ and $\mathbf{Y}_i = \{\mathbf{y}_i^j\}_{j=1}^{N_i}$, respectively, where N_i is the number of source samples. Let $\mathbf{X}_T = \{\mathbf{x}_T^j\}_{j=1}^{N_T}$ and $\mathbf{Y}_T = \{\mathbf{y}_T^j\}_{j=1}^{N_T}$ denote the target data and corresponding labels drawn from the target distribution $P_T(\mathbf{x}, \mathbf{y})$, where N_T is the number of target samples.

Suppose the number of labeled target samples is N_{TL}, the MDA problem can be classified into different categories:

- *unsupervised MDA*, when $N_{TL} = 0$;
- *fully supervised MDA*, when $N_{TL} = N_T$;
- *semi-supervised MDA*, otherwise.

Suppose $\mathbf{x}_i^j \in \mathbb{R}^{d_i}$ and $\mathbf{x}_T^j \in \mathbb{R}^{d_T}$ are an observation in S_i and T, respectively, we can classify MDA into:

- *homogeneous MDA*, when $d_1 = \cdots = d_M = d_T$;
- *heterogeneous MDA*, otherwise.

Suppose $\mathcal{C}_i$ and $\mathcal{C}_T$ are the label set for source S_i and target T, we can define different MDA strategies:

- *closed-set MDA*, when $\mathcal{C}_1 = \cdots = \mathcal{C}_M = \mathcal{C}_T$;
- *open-set MDA*, for all $\mathcal{C}_i, \mathcal{C}_i \cap \mathcal{C}_T \subset \mathcal{C}_T$;
- *partial MDA*, when $\mathcal{C}_T \subset (\mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots \cup \mathcal{C}_M)$;
- *union-set MDA*, $\mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots \cup \mathcal{C}_M = \mathcal{C}_T$;
- *universal MDA*, without prior knowledge on label sets;

where $\cap$, $\cup$, and $\subset$ respectively indicate the intersection set, union set, and proper subset between two sets.

Suppose the number of labeled source samples is N_{iL} for source S_i, the MDA problem can be classified into:

- *strongly supervised MDA*, when $N_{iL} = N_i$ for all i;
- *few-shot MDA*, when $N_{iL} << N_i$ for all i.

Based on the data privacy and security constraints, the MDA task can be classified into:

- *centralized MDA*, when $\mathbf{X}_i(i = 1, \cdots, M)$ and $\mathbf{X}_T$ are available and shared during adaptation;
- *federated MDA*, when $\mathbf{X}_i(i = 1, \cdots, M)$ and $\mathbf{X}_T$ are available but unshared during adaptation;
- *source-free MDA*, the trained source models are available but the source data cannot be accessed.

Multi-target MDA refers to adapting to multiple target domains simultaneously. When the target data is unavailable during training, the task becomes domain generalization.

3 Deep MDA Taxonomy

Existing methods on deep MDA primarily focus on the unsupervised, homogeneous, closed-set, strongly supervised, centralized, one target, and target data available settings. That is, there is one target domain, the target data is unlabeled but available during adaptation, the source data is fully labeled and available during adaptation, the source and target data are observed in the same data space, and the label sets of all sources and the target are the same. In this paper, we will focus on MDA methods under these settings and will also briefly introduce other specific MDA settings.

There is some theoretical analysis to support existing MDA algorithms based on different statistical learning frameworks, such as VC dimension, Rademacher complexity, information theory, etc. Most theories are based on the seminal theoretical model [Blitzer *et al.*, 2008; Ben-David *et al.*, 2010], with an upper bound derived from VC dimension. Mansour *et al.* [2009] assumed that the target distribution can be approximated by a convex combination of the M source distributions and provided a distribution-weighted hypothesis combination rule. Moreover, tighter cross-domain generalization bounds and more accurate measurements of domain discrepancy can provide intuitions to derive effective MDA algorithms. Hoffman *et al.* [2018] derived a novel bound using DC-decomposition and calculated more accurate combination weights of the non-convex optimization situation. Zhao *et al.* [2018] extended the generalization bound of the seminal theoretical model to multiple sources under both classification and regression settings. Besides the domain discrepancy between the target and each source [Hoffman *et al.*, 2018; Zhao *et al.*, 2018], Li *et al.* [2018b] also considered the relationship between pairwise sources and derived a tighter bound on the weighted multi-source discrepancy. Based on this bound, more relevant source domains can be picked out. A finite-sample error bound is derived [Wen *et al.*, 2020] using sample-based Rademacher complexity, a generally tighter complexity measure than VC dimension. The attention mechanism is proven to automatically reduce the negative effects caused by unrelated domains [Zuo *et al.*, 2021].

The label distribution has also been theoretically studied. To deal with the inconsistency between the source and target label distributions, Redko *et al.* [2019] provided a generalization bound when transferring under target shift. Later, the similarity of the semantic conditional distribution is designed to aggregate multiple sources [Shui *et al.*, 2021]. Mutual information in the information theory provides another theoretical solution. Park and Lee [2021] proved that the mutual information between latent representations and domain labels is closely related to the explicit optimization of the divergence between the source and target domains. Recently, Chen and Marchand [2023] provided algorithm-dependent generalization bounds characterized by the mutual information between the parameters and the data.

Typically, some task models (*e.g.*, classifiers) are learned based on the labeled source data with corresponding task loss. Meanwhile, specific alignments among the source and target domains are conducted to bridge the domain shift so that the learned task classifiers can be better generalized to the target domain. Based on the alignment methodologies on different levels, we can classify MDA into different categories, as shown in Figure 3. *Latent space transformation* that performs feature-level alignment tries to learn domain-invariant features of different domains. *Intermediate domain generation* aims to align different domains on the data pixel level by explicitly generating an intermediate adapted domain for each source that is indistinguishable from the target domain. *Task classifier refinement* considers the label distribution variance so that the learned task classifiers can better fit the target domain. Latent space transformation and intermediate domain generation focus on bridging the covariate shift, *i.e.*, the con-
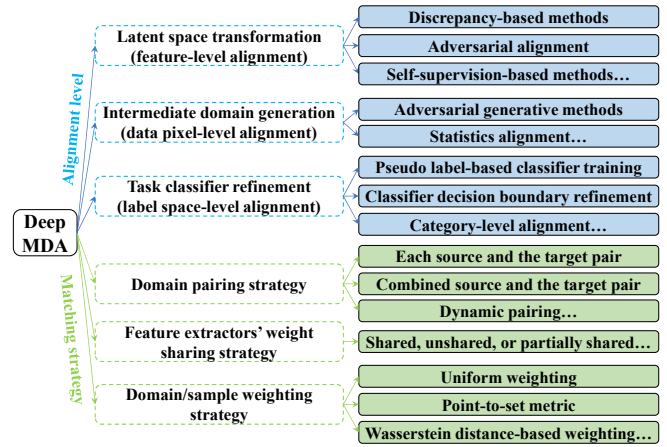

Figure 3: Taxonomy of deep MDA methods.

ditional distributions of labels given an observation are the same across domains, but the marginal distributions of observations are different. Task classifier refinement aims to address the label shift where the marginal label distributions differ across domains. Figure 4 illustrates a widely employed framework of existing MDA methods. Meanwhile, several matching strategies should be considered to deal with the case of multiple sources and one target, including domain pairing strategy, feature extractors' weight sharing strategy, and domain/sample weighting strategy.

4 MDA Alignment Methodologies

4.1 Latent Space Transformation

The common methods for aligning the latent spaces of different domains, *i.e.*, learning domain-invariant feature representations, include discrepancy-based methods, adversarial alignment, and self-supervision-based methods.

Discrepancy-based methods measure the discrepancy of the latent spaces (typically features) from different domains by optimizing some specific discrepancy losses, such as maximum mean discrepancy (MMD) [Guo *et al.*, 2018; Zhu *et al.*, 2019; Li *et al.*, 2022; Wang *et al.*, 2022], Rényi-divergence [Hoffman *et al.*, 2018], moment distance [Peng *et al.*, 2019; Fu *et al.*, 2021], weighted moment distance [Zuo *et al.*, 2021], correlation alignment [Karisani, 2022], relation alignment loss [Wang *et al.*, 2020], contrastive domain discrepancy [Kang *et al.*, 2022], $\mathcal{L}_1$ distance [Cai *et al.*, 2023], and Wasserstein distance [Chen and Marchand, 2023]. Guo *et al.* [2020] claimed that different discrepancies or distances can only provide specific estimates of domain similarities and thus they considered the mixture of multiple distances. Minimizing the discrepancy to align the features among the source and target domains usually does not introduce any new to-be-learned parameters. Instead of using the above-mentioned explicit discrepancy metrics, Li *et al.* [2018a] proposed Adaptive Batch Normalization (AdaBN) to implicitly minimize the discrepancy between the source and target domains by aligning the BN statistics, such as the moving average mean and variance of all BN layers.
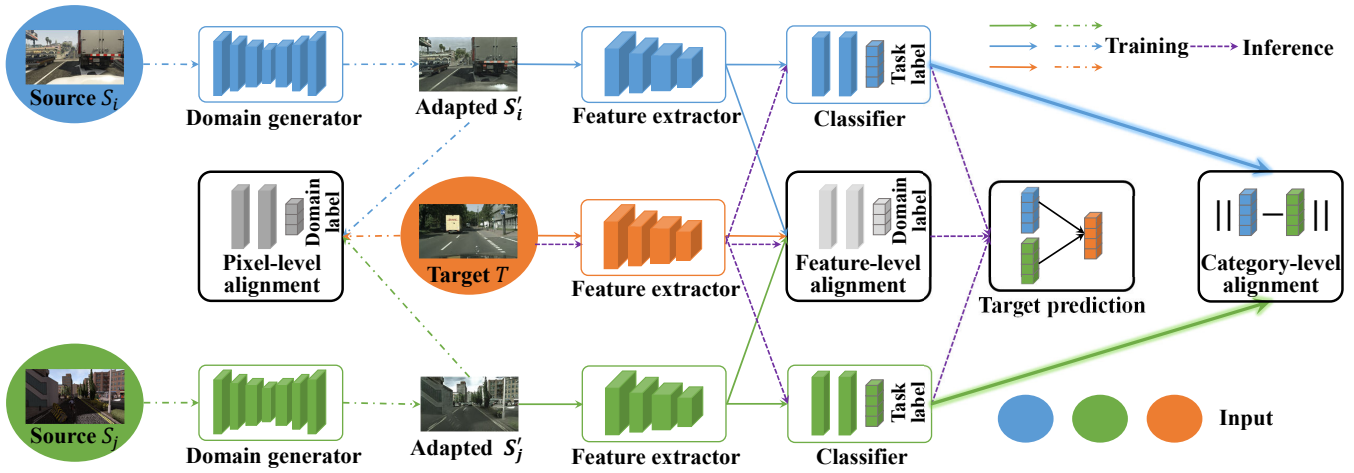
Figure 4: A widely employed MDA framework. The solid arrows, dashed dot arrows, and bold arrows indicate the training of latent space transformation, intermediate domain generation, and task classifier refinement, respectively. The dashed arrows indicate the reference process.

Adversarial alignment employs a discriminator to align the features by making them indistinguishable. Some representative optimized objectives include GAN loss [Xu *et al.*, 2018; Zhao *et al.*, 2021a; Nguyen *et al.*, 2021; Liu and Ren, 2022; Zhang *et al.*, 2022], weighted GAN loss [Yang *et al.*, 2020; Yin *et al.*, 2022], attention-guided GAN loss [Gong *et al.*, 2021], $\mathcal{H}$-divergence [Zhao *et al.*, 2018; Park and Lee, 2021], $d_{\mathcal{A}}$ distance [Wen *et al.*, 2020], Wasserstein distance [Li *et al.*, 2018b; Zhao *et al.*, 2020], least squares loss and focal loss [Yao *et al.*, 2021a], weighted conditional Wasserstein distance [Shui *et al.*, 2021], cross-entropy loss [Wu and Shi, 2022; Zhou *et al.*, 2022; Belal *et al.*, 2024], $\mathcal{L}_2$ 0loss [Yao *et al.*, 2021b], and metric constraint loss [Ren *et al.*, 2022]. These methods aim to confuse the discriminator's ability to distinguish whether the features from multiple sources were drawn from the same distribution. Compared with GAN loss and $\mathcal{H}$-divergence, Wasserstein distance can provide more stable gradients even when the target and source distributions do not overlap [Zhao *et al.*, 2020]. Instead of directly taking the output of feature extractors as input into the discriminator, Zhao *et al.* [2021b] employed another generator to generate adapted representations, which are input into the discriminator for alignment. The discriminator is often implemented as a network, which leads to new parameters that need to be learned.

Self-supervision-based methods typically combine some auxiliary self-supervised learning (pretext) tasks with the target task. Some commonly used pretext tasks include reconstruction, image rotation, and patch location prediction [Zhao *et al.*, 2022]. In [Yuan *et al.*, 2022], the domain information is considered as the labels of the pretext task. That is, domain classification plays the role of pretext task.

4.2 Intermediate Domain Generation

Feature-level alignment only aligns high-level information, which is insufficient for fine-grained predictions, such as pixel-wise semantic segmentation [Zhao *et al.*, 2019; Zhao *et al.*, 2021a]. Generating an intermediate adapted domain with pixel-level alignment, typically via GAN, diffusion models, and their variants, can help address this problem. Zhao *et al.* [2021a] used CycleGAN to generate an adapted domain for each source and proposed to aggregate different adapted domains using a sub-domain aggregation discriminator and a cross-domain cycle discriminator, where the pixel-level alignment is then conducted between the aggregated and target domains. Instead of taking the original source data as input to the generator, Lin *et al.* [2020] used a variational autoencoder to map all source and target domains to a latent space and then generated an adapted domain from the latent space. Zhao *et al.* [2021a] and Lin *et al.* [2020] showed that the semantics might change in the intermediate domain and that enforcing a semantic consistency before and after generation can help preserve the original semantics. He *et al.* [2021] proposed to translate the style of images in source domains to the style of the target domain by aligning the statistics distribution of pixel values in LAB color space. Besides the standard GAN loss, Kang *et al.* [2023] further enforced the minimization of a mutual information loss and a texture co-occurrence loss to respectively constrain structural changes and encourage high-quality texture translation in local areas.

4.3 Task Classifier Refinement

Even after the source and target domains are aligned on the feature level and pixel level by latent space transformation and intermediate domain generation, the learned task classifiers on the labeled sources still cannot be guaranteed to perform well on the target. For example, the classifiers trained on different sources may result in misaligned predictions for the target samples that are close to the domain boundary; the class distributions are imbalanced across domains. Task classifier refinement aims to address these issues.

One intuitive method is **pseudo label-based classifier training** which employs the target pseudo labels for classifier training [Venkat *et al.*, 2020; He *et al.*, 2021]. A logit-based metric is designed to select reliable pseudo labels [Venkat *et al.*, 2020]. **Classifier decision boundary re-**

finement tries to refine the decision boundary of the classifier to better fit the target domain. Specific classifier discrepancies, such as $\mathcal{L}_1$ loss [Zhu *et al.*, 2019; Peng *et al.*, 2019; Yao *et al.*, 2021a], and divergence loss [Nguyen *et al.*, 2021; Karisani, 2022], are minimized to encourage the predictions of the same target sample by different classifiers to be the same. Both intra-domain and inter-domain consistency losses based on $\mathcal{L}_1$ are considered [Luo *et al.*, 2022]. Boundary-sensitive alignment is proposed to assign different weights to the source-target alignments based on the cross-entropy loss of the labeled source samples [Kang *et al.*, 2022]. The dynamic classifier alignment method aligns classifiers derived from multi-view features through an automatic, sample-wise approach [Li *et al.*, 2023a].

Category-level alignment aligns the source and target domains, usually based on the feature representations, for each category. MADAN+ employs one discriminator for each category to align the features of different domains in an adversarial manner [Zhao *et al.*, 2021a]. Differently, category-level discrepancy-based alignment incorporates the class distribution prior information when computing the MMD discrepancy between different domains [Wang *et al.*, 2022; Li *et al.*, 2022]. Since the target domain is unlabeled, obtaining reliable pseudo target labels plays an important role in category-level alignment. Categorical prototypes are aligned across different domains based on minimax entropy loss [Liu and Ren, 2022], reconstruction [Zhou *et al.*, 2022], and cosine similarity [Belal *et al.*, 2024]. The SIG model integrates class-aware conditional alignment in domain-invariant and label-relevant latent variables to address changes in label distributions across different domains [Li *et al.*, 2023b].

5 MDA Matching Strategies

5.1 Domain Pairing Strategy

There are different ways to group the target and multiple sources so that alignment can be performed. The most common method is to pairwise align the target with each source, such as [Xu *et al.*, 2018; Zhao *et al.*, 2018; Zhao *et al.*, 2020; Karisani, 2022], to name a few. Since domain shift also exists among different sources, several methods also enforce pairwise alignment between all sources [Li *et al.*, 2018b; Peng *et al.*, 2019; Yin *et al.*, 2022]. Different adapted domains are first aggregated into one domain, which is then aligned with the target [Zhao *et al.*, 2021a; Li *et al.*, 2021b; Venkat *et al.*, 2020; Chen and Marchand, 2023]. A multi-armed bandit controller is employed to dynamically select one source to learn an optimal trajectory and mixture of domains for adaptation [Guo *et al.*, 2020]. All the source and target domains are directly discriminated by a multi-class discriminator [Park and Lee, 2021]. A tensor-low-rank constraint is imposed on the prototypical similarity matrices to explore the high-order relationships among different domains [Li *et al.*, 2021a].

All the above methods assume that the domain label information is available. However, this might not hold in practice. In such cases, a naive approach is to directly combine all the sources into one to align with the target [Yang *et al.*, 2020; Zhao *et al.*, 2021b; Li *et al.*, 2021b].

5.2 Feature Extractors' Weight Sharing Strategy

Most MDA methods employ shared feature extractors to learn domain-invariant features, such as [Zhao *et al.*, 2018; Peng *et al.*, 2019; Kang *et al.*, 2022; Wang *et al.*, 2022; Li *et al.*, 2023b; Belal *et al.*, 2024]. However, domain invariance may be detrimental to the discriminative power and it might be difficult to extract the same domain-invariant features for all domains. Besides the same feature extractor, M domain-specific feature extractors are trained so that each source and the target can be better aligned [Zhu *et al.*, 2019], while only M domain-specific feature extractors are considered in [He *et al.*, 2021]. One feature extractor is first pre-trained for each source and then the target is mapped into the feature space of each source [Zhao *et al.*, 2020]. The source domains share a feature extractor and the target uses a separate one [Wu *et al.*, 2022]. The source domains with similar characteristics based on coordination share the same feature extractor [Karisani, 2022]. Low-level features that are important for localization share the same extractor across domains with strong alignment, while high-level features that are important for object recognition use specific extractors with weak alignment [Yao *et al.*, 2021a; Zhang *et al.*, 2022]. Although unshared feature extractors can better align the target and sources in the latent space, this substantially increases the number of parameters in the model.

5.3 Domain/Sample Weighting Strategy

Some early deep MDA methods treat different source domains and source samples equally. For example, only one task classifier is trained for all the source domains even with available domain label information [Zhao *et al.*, 2018]. However, theis ignores the different correlations between multiple sources and the target as well as the different similarities between source and target samples. Assigning different source domains and source samples with optimal weights would result in better alignment and generalization. One common domain weighting strategy is to combine the target predictions by multiple source classifiers with different weights during inference, such as uniform weight [Zhu *et al.*, 2019], perplexity score based on adversarial loss [Xu *et al.*, 2018], point-to-set (PoS) metric using Mahalanobis distance [Guo *et al.*, 2018], relative error based on source-only accuracy [Peng *et al.*, 2019], Wasserstein distance based weights [Zhao *et al.*, 2020], and majority voting [Karisani, 2022].

The other domain weighting strategy aims to adjust the importance weights of source domains during training [Wen *et al.*, 2020]. An uncertainty-aware weighting strategy is designed to adaptively assign weights to different source domains and samples [Li *et al.*, 2021a]. Based on the attention- or discriminator-based domain correlations that measure the probabilities of target samples belonging to each source domain, weighted moment distance is designed to pay more attention to the source domains with higher similarities during alignment [Zuo *et al.*, 2021; Yao *et al.*, 2021a; Li *et al.*, 2023a]. Domain weights are derived based on the upper bound or the gradient norm bound [Shui *et al.*, 2021; Chen and Marchand, 2023].

Sample weighting has also been explored recently. Based on a manually selected Wasserstein distance threshold, the

source samples that are closer to the target samples are distilled to fine-tune the source classifiers [Zhao *et al.*, 2020]. During adversarial alignment, the source samples in a batch are assigned with different weights by model-based curriculum with a source selection network [Yang *et al.*, 2020; Zhao *et al.*, 2021b] or by model-free curriculum based on the output of domain discriminators [Zhao *et al.*, 2021b; Liu and Ren, 2022; Zhang *et al.*, 2022]. The source samples are re-weighted based on the source-target relevance obtained by the nearest neighbor algorithm [Wu *et al.*, 2022]. A weighted loss term with a general self-paced regularizer is employed to optimize both model parameters and sample weights [Wang *et al.*, 2022]. The source features are weighted by target similarity evaluated by the normalized domain discriminator output [Liu and Ren, 2022].

6 Other Specific MDA Settings

Heterogeneous MDA. Transforming the heterogeneous input from different domains into a shared latent space with the same feature dimensions is an intuitive solution. The feature extractor for each domain contains two networks [Yao *et al.*, 2021b]. One is domain-specific with unshared structures and parameters. The other has identical structures with a $\mathcal{L}_1$-based parameter consistency constraint, which aims to make the parameters consistent across domains.

Open-set, partial, and universal MDA. The core of open-set MDA lies in how to align the known classes between the sources and the target while making the unknown classes separable. Rakshit *et al.* [2020] studied the case that all sources share the same label set, while the target contains unknown classes. A binary pseudo classifier with a margin loss is trained to classify whether a target sample belongs to the unknown class. In partial MDA, the relevant samples in the shared classes need to be emphasized while the irrelevant ones in the sources need to be ignored. In [Fu *et al.*, 2021], a selection vector is first derived to refine the extracted features and then the selected features are aligned on both the domain level and the category level using $\mathcal{L}_2$ loss and local relation alignment Loss, respectively. An attention-guided adversarial alignment and pseudo-label supervision fusion mechanisms are designed to effectively harness both labeled and unlabeled data across domains [Gong *et al.*, 2021]. The main challenge in universal MDA is to identify the common label set among each source and the target and keep the model scalable as the number of source domains increases. A pseudo-margin vector that measures the confidence in assigning a target sample to the pseudo label helps to identify the common label set and to adversarially align the distributions of multiple sources and the unknown target [Yin *et al.*, 2022].

Federated and source-free MDA. In the pursuit of privacy protection, decentralized MDA, such as federated and source-free MDA, has emerged. In federated MDA, the model parameters at each node are trained separately with convergence at different rates. Peng *et al.* [2020] first introduced federated adversarial adaptation to optimize the $\mathcal{H}$-divergence without accessing data, while employing a dynamic attention mechanism and leveraging representation disentanglement to mitigate negative transfer to the target domain. To reduce the communication costs and prevent privacy leakage attacks, Feng *et al.* [2021] proposed a multi-source knowledge distillation method and a dynamic weighting strategy to obtain high-quality domain consensus and identify the irrelevant sources. Based on whether the parameters of pre-trained source models are available, source-free MDA (SFMDA) can be divided into white box and black box. A common white box SFMDA pipeline involves using pseudo labels, which often introduce noise that impacts the adaptation result. Several pseudo-label denoising strategies have been proposed, such as pseudo label selection based on self-supervised clustering [Ahmed *et al.*, 2021] and confidence query [Shen *et al.*, 2023], confident-anchor-induced pseudo label generation [Dong *et al.*, 2021], and multi-source probabilistic ensemble [Zhao *et al.*, 2024]. A source importance estimation module [Han *et al.*, 2023] and a concurrent subsidiary supervision module [Kundu *et al.*, 2022] are designed to integrate into existing models to enhance the adaptation performance. For black box SFMDA, the knowledge from source predictors is first distilled to a customized target model, which is then fine-tuned with the target data [Liang *et al.*, 2022].

Meta-learning-based MDA. Meta-learning aims to learn a general model from multiple tasks by induction. Incorporating meta-learning into MDA involves creating algorithms that can discern and leverage the underlying structure shared across different domains. Li and Hospedales [Li and Hospedales, 2020] proposed to use meta-learning to enhance the performance of existing MDA methods. Specifically, the multi-source domains are split into disjoint meta-training and meta-testing domains. The best initial condition that is suited to adapting between all source domains is searched for the sake of adapting well to the target domain.

7 Applications, Datasets, and Benchmark

MDA has been successfully applied in various areas, ranging from computer vision (CV) to natural language processing (NLP) and from speech recognition to industry diagnosis. Specifically, MDA has been studied in multiple learning tasks, such as image classification, object detection, and semantic segmentation in CV, as well as sentiment classification and machine translation in NLP. The datasets for evaluating MDA models usually contain multiple domains with different styles, such as *synthetic* vs. *real*, *artistic* vs. *sketchy*, which impose large domain shift. Here we briefly introduce some commonly used datasets in both CV and NLP areas.

Image classification is the most popular CV task for MDA. Some representative datasets include Digits-five for digit recognition, Office-31, Office-Caltech, ImageCLEF, PACS, Office-Home, and DomainNet for object classification. Existing MDA methods can obtain an accuracy that is higher than 90% for some datasets (*e.g.*, Digits-five, Office-31, Office-Caltech, and ImageCLEF) and even 100% for some domains (*e.g.*, Webcam and DSLR in Office-Caltech) [Yuan *et al.*, 2022]. Office-Home and DomainNet [Peng *et al.*, 2019] are more challenging. Office-Home consists of about 15,500 images from 65 categories of everyday objects in office and home settings with 4 domains. DomainNet contains about 600K images from 345 categories with 6 domains: Clipart

	Clp	Inf	Pnt	Qdr	Rel	Skt	Avg
Source-B	39.6	8.2	33.9	11.8	41.6	23.1	26.4
Source-C	47.6	13.0	38.1	13.3	51.9	33.7	32.9
DCTN	48.6	23.5	48.8	7.2	53.5	47.3	38.2
MDAN	52.4	21.3	46.9	8.6	54.9	46.5	38.4
M3SDA	58.6	26.0	52.3	6.3	62.7	49.5	42.6
MDDA	59.4	23.8	53.2	12.5	61.8	48.6	43.2
CMSS	64.2	28.0	53.6	16.0	63.4	53.8	46.5
LtC-MSDA	63.1	28.7	56.1	16.3	66.1	53.8	47.4
SImPAI	66.4	26.5	56.6	18.9	68.0	55.5	48.6
T-SVDNet	66.1	25.0	54.3	16.5	65.4	54.6	47.0
ABMSDA	66.9	25.3	55.8	18.2	64.1	55.2	47.6
DEAL	70.8	26.5	57.4	12.2	65.0	60.6	48.7
DRT+ST	71.0	31.6	61.0	12.3	71.4	**60.7**	51.3
PTMDA	66.0	28.5	58.4	13.0	63.0	54.1	47.2
SPS	70.8	24.6	55.2	19.4	67.5	57.6	49.2
SSG	68.7	24.8	55.7	18.4	68.8	56.3	48.8
MSCAN	69.3	28.0	58.6	**30.3**	**73.3**	59.5	**53.2**
SIG	**72.7**	**32.0**	**61.5**	20.5	72.4	59.5	53.0
Oracle	69.3	34.5	66.3	66.8	80.1	60.7	63.0

Table 1: Classification accuracy (%) of different adaptation tasks on DomainNet based on ResNet-101. The best performance trained on the source domains is emphasized in bold.

(*Clp*), Infograph (*Inf*), Painting (*Pnt*), Quickdraw (*Qdr*), Real (*Rel*), and Sketch (*Skt*). Sim2Real MDA is often performed for object detection and semantic segmentation. Commonly used synthetic domains include GTA and SYNTHIA, while Cityscapes, KITTI, and BDD100K are popular real domains.

Sentiment classification is a widely-studied NLP task for MDA. Some representative datasets include Amazon Reviews, Media Reviews, Reviews-5, and Multilingual Amazon Reviews Corpus (MARC). As a relatively new dataset, MARC is a collection of Amazon reviews from four languages: German, English, French, and Japanese. For each language, there are three domains: Books, DVD, and Music. We can perform cross-domain adaptation for each language and cross-lingual adaptation for each domain.

To give readers a brief and quick understanding of how current MDA methods perform, we report the object classification accuracy of several adaptation tasks on DomainNet. One domain is selected as the target and the rest domains are used as multiple sources. The compared MDA methods include DCTN [Xu *et al.*, 2018], MDAN [Zhao *et al.*, 2018], M3SDA [Peng *et al.*, 2019], MDDA [Zhao *et al.*, 2020], CMSS [Yang *et al.*, 2020], LtC-MSDA [Wang *et al.*, 2020], SImPAI [Venkat *et al.*, 2020], T-SVDNet [Li *et al.*, 2021a], ABMSDA [Zuo *et al.*, 2021], DEAL [Zhou *et al.*, 2022], DRT+ST [Li *et al.*, 2021b], PTMDA [Ren *et al.*, 2022], SPS [Wang *et al.*, 2022], SSG [Yuan *et al.*, 2022], MSCAN [Kang *et al.*, 2022], and SIG [Li *et al.*, 2023b]. For comparison, we report the results of source-only methods that are trained on the source domains and directly tested on the target domain under two settings: single-best ('Source-B') and source-combined ('Source-C'). The result of an oracle setting is also reported, where the classification model is both trained and tested on the target domain. The results of source-only and oracle methods can be viewed as a lower bound and an upper bound of MDA.

From the results in Table 1, we have the following observations: (1) Because of the domain gap, the source-only methods without any alignment perform the worst in most adaptation settings. (2) Through alignment on different levels, such as the feature level, the MDA methods can significantly improve the classification accuracy. As compared to Source-C, MSCAN achieves a 20.3% performance gain. (3) For specific domains, such as Clp, some adaptation algorithms (*e.g.*, SIG) can even outperform the oracle method, demonstrating the effectiveness and necessity of MDA. (4) On average, there is still a large performance gap between MDA methods and oracle, which calls for further research efforts to better reduce the domain gap.

8 Conclusion and Future Directions

In this paper, we provided a systematic survey of recent MDA developments in the deep learning era. We motivated MDA, defined different strategies, and summarized the commonly used datasets with a brief benchmark. We classified existing methods into different categories and compared the representative ones technically and experimentally. We conclude this paper with several open research directions:

Specific MDA strategy implementation. There are many MDA strategies defined from different aspects. In practice, we might encounter specific combinations of these strategies, such as union-set source-free MDA. Specific implementation of such an MDA strategy would likely yield better results than a one-size-fits-all MDA approach.

Multi-modal and time-series MDA. The labeled source data may be of different modalities, such as LiDAR, radar, and image. Further research is needed to find techniques for fusing different data modalities in MDA. A further extension of this idea is to have varied modalities in different sources and the modalities contain time-series information.

Incremental and test-time MDA. Incremental MDA remains largely unexplored and may provide great benefits for real-world scenarios, such as updating deployed MDA models when new source or target data becomes available. Test-time MDA that allows more flexible inference time can provide inference unlabeled data for adaptation.

Large-scale pre-training and self-learning for MDA. The rapid development of large-scale pre-training and self-learning has been witnessed recently, such as BERT and GPT. Employing such pre-training and self-learning techniques to explore their general representation for MDA is a promising direction.

Theoretical and interpretable analysis. Theoretical and interpretable analysis can guide the design of MDA algorithms and provide novel insights to understand the pros and cons of MDA models. For example, we may derive a tighter upper bound, analyze whether every source domain and source sample contributes, and visualize the learned features to see if different categories are correctly classified.

Acknowledgements

References

[Ahmed *et al.*, 2021] Sk Miraj Ahmed, Dripta S Raychaudhuri, et al. Unsupervised multi-source domain adaptation without access to source data. In *CVPR*, 2021.

[Belal *et al.*, 2024] Atif Belal, Akhil Meethal, et al. Multi-source domain adaptation for object detection with prototype-based mean teacher. In *WACV*, 2024.

[Ben-David *et al.*, 2010] Shai Ben-David, John Blitzer, et al. A theory of learning from different domains. *ML*, 2010.

[Blitzer *et al.*, 2008] John Blitzer, Koby Crammer, et al. Learning bounds for domain adaptation. In *NIPS*, 2008.

[Cai *et al.*, 2023] Yuxiang Cai, Meng Xi, Yongheng Shang, and Jianwei Yin. Exploring high-correlation source domain information for multi-source domain adaptation in semantic segmentation. In *ACM MM*, 2023.

[Chen and Marchand, 2023] Qi Chen and Mario Marchand. Algorithm-dependent bounds for representation learning of multi-source domain adaptation. In *AISTATS*, 2023.

[Dong *et al.*, 2021] Jiahua Dong, Zhen Fang, Anjin Liu, Gan Sun, and Tongliang Liu. Confident anchor-induced multi-source free domain adaptation. In *NeurIPS*, 2021.

[Feng *et al.*, 2021] Haozhe Feng, Zhaoyang You, et al. Kd3a: Unsupervised multi-source decentralized domain adaptation via knowledge distillation. In *ICML*, 2021.

[Fu *et al.*, 2021] Yangye Fu, Ming Zhang, et al. Partial feature selection and alignment for multi-source domain adaptation. In *CVPR*, 2021.

[Ganin *et al.*, 2016] Yaroslav Ganin, Evgeniya Ustinova, et al. Domain-adversarial training of neural networks. *JMLR*, 2016.

[Gong *et al.*, 2021] Rui Gong, Dengxin Dai, et al. mdalu: Multi-source domain adaptation and label unification with partial datasets. In *ICCV*, 2021.

[Guo *et al.*, 2018] Jiang Guo, Darsh Shah, and Regina Barzilay. Multi-source domain adaptation with mixture of experts. In *EMNLP*, 2018.

[Guo *et al.*, 2020] Han Guo, Ramakanth Pasunuru, and Mohit Bansal. Multi-source domain adaptation for text classification via distancenet-bandits. In *AAAI*, 2020.

[Han *et al.*, 2023] Zhongyi Han, Zhiyan Zhang, et al. Discriminability and transferability estimation: a bayesian source importance estimation approach for multi-source-free domain adaptation. In *AAAI*, 2023.

[He *et al.*, 2021] Jianzhong He, Xu Jia, et al. Multi-source domain adaptation with collaborative learning for semantic segmentation. In *CVPR*, 2021.

[Hoffman *et al.*, 2018] Judy Hoffman, Mehryar Mohri, and Ningshan Zhang. Algorithms and theory for multiple-source adaptation. In *NIPS*, 2018.

[Kang *et al.*, 2022] Guoliang Kang, Lu Jiang, et al. Contrastive adaptation network for single-and multi-source domain adaptation. *IEEE TPAMI*, 2022.

[Kang *et al.*, 2023] Myeongkyun Kang, Philip Chikontwe, et al. Structure-preserving image translation for multi-source medical image domain adaptation. *PR*, 2023.

[Karisani, 2022] Payam Karisani. Multiple-source domain adaptation via coordinated domain encoders and paired classifiers. In *AAAI*, 2022.

[Kundu *et al.*, 2022] Jogendra Nath Kundu, Suvaansh Bhambri, et al. Concurrent subsidiary supervision for unsupervised source-free domain adaptation. In *ECCV*, 2022.

[Li and Hospedales, 2020] Da Li and Timothy Hospedales. Online meta-learning for multi-source and semi-supervised domain adaptation. In *ECCV*, 2020.

[Li *et al.*, 2018a] Yanghao Li, Naiyan Wang, Jianping Shi, Xiaodi Hou, and Jiaying Liu. Adaptive batch normalization for practical domain adaptation. *PR*, 2018.

[Li *et al.*, 2018b] Yitong Li, Michael Andrew Murias, et al. Extracting relationships by multi-domain matching. In *NIPS*, 2018.

[Li *et al.*, 2021a] Ruihuang Li, Xu Jia, et al. T-svdnet: Exploring high-order prototypical correlations for multi-source domain adaptation. In *ICCV*, 2021.

[Li *et al.*, 2021b] Yunsheng Li, Lu Yuan, Yinpeng Chen, Pei Wang, and Nuno Vasconcelos. Dynamic transfer for multi-source domain adaptation. In *CVPR*, 2021.

[Li *et al.*, 2022] Keqiuyin Li, Jie Lu, Hua Zuo, and Guangquan Zhang. Multi-source contribution learning for domain adaptation. *IEEE TNNLS*, 2022.

[Li *et al.*, 2023a] Keqiuyin Li, Jie Lu, et al. Dynamic classifier alignment for unsupervised multi-source domain adaptation. *IEEE TKDE*, 2023.

[Li *et al.*, 2023b] Zijian Li, Ruichu Cai, et al. Subspace identification for multi-source domain adaptation. In *NeurIPS*, 2023.

[Liang *et al.*, 2022] Jian Liang, Dapeng Hu, Jiashi Feng, and Ran He. Dine: Domain adaptation from single and multiple black-box predictors. In *CVPR*, 2022.

[Lin *et al.*, 2020] Chuang Lin, Sicheng Zhao, Lei Meng, and Tat-Seng Chua. Multi-source domain adaptation for visual sentiment classification. In *AAAI*, 2020.

[Liu and Ren, 2022] Yong-Hui Liu and Chuan-Xian Ren. A two-way alignment approach for unsupervised multi-source domain adaptation. *PR*, 2022.

[Luo *et al.*, 2022] Zhipeng Luo, Xiaobing Zhang, et al. Domain consistency regularization for unsupervised multi-source domain adaptive classification. *PR*, 2022.

[Mansour *et al.*, 2009] Yishay Mansour, Mehryar Mohri, and Afshin. Rostamizadeh. Domain adaptation with multiple sources. In *NIPS*, 2009.

[Nguyen *et al.*, 2021] Van-Anh Nguyen, Tuan Nguyen, et al. Stem: An approach to multi-source domain adaptation with guarantees. In *ICCV*, 2021.

[Park and Lee, 2021] Geon Yeong Park and Sang Wan Lee. Information-theoretic regularization for multi-source domain adaptation. In *ICCV*, 2021.

[Peng *et al.*, 2019] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *ICCV*, 2019.

[Peng *et al.*, 2020] Xingchao Peng, Zijun Huang, Yizhe Zhu, and Kate Saenko. Federated adversarial domain adaptation. In *ICLR*, 2020.

[Rakshit *et al.*, 2020] Sayan Rakshit, Dipesh Tamboli, et al. Multi-source open-set deep adversarial domain adaptation. In *ECCV*, 2020.

[Redko *et al.*, 2019] Ievgen Redko, Nicolas Courty, et al. Optimal transport for multi-source domain adaptation under target shift. In *AISTATS*, 2019.

[Ren *et al.*, 2022] Chuan-Xian Ren, Yong-Hui Liu, et al. Multi-source unsupervised domain adaptation via pseudo target domain. *IEEE TIP*, 2022.

[Shen *et al.*, 2023] Maohao Shen, Yuheng Bu, and Gregory W Wornell. On balancing bias and variance in unsupervised multi-source-free domain adaptation. In *ICML*, 2023.

[Shui *et al.*, 2021] Changjian Shui, Zijian Li, Jiaqi Li, Christian Gagné, Charles X Ling, and Boyu Wang. Aggregating from multiple target-shifted sources. In *ICML*, 2021.

[Sun *et al.*, 2015] Shiliang Sun, Honglei Shi, and Yuanbin Wu. A survey of multi-source domain adaptation. *INFFUS*, 2015.

[Venkat *et al.*, 2020] Naveen Venkat, Jogendra Nath Kundu, et al. Your classifier can secretly suffice multi-source domain adaptation. In *NeurIPS*, 2020.

[Wang *et al.*, 2020] Hang Wang, Minghao Xu, et al. Learning to combine: Knowledge aggregation for multi-source domain adaptation. In *ECCV*, 2020.

[Wang *et al.*, 2022] Zengmao Wang, Chaoyang Zhou, Bo Du, and Fengxiang He. Self-paced supervision for multi-source domain adaptation. In *IJCAI*, 2022.

[Wen *et al.*, 2020] Junfeng Wen, Russell Greiner, and Dale Schuurmans. Domain aggregation networks for multi-source domain adaptation. In *ICML*, 2020.

[Wilson and Cook, 2020] Garrett Wilson and Diane J Cook. A survey of unsupervised deep domain adaptation. *ACM TIST*, 2020.

[Wu and Shi, 2022] Hui Wu and Xiaodong Shi. Adversarial soft prompt tuning for cross-domain sentiment analysis. In *ACL*, 2022.

[Wu *et al.*, 2022] Jiaxi Wu, Jiaxin Chen, et al. Target-relevant knowledge preservation for multi-source domain adaptive object detection. In *CVPR*, 2022.

[Xu *et al.*, 2018] Ruijia Xu, Ziliang Chen, et al. Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In *CVPR*, 2018.

[Yang *et al.*, 2020] Luyu Yang, Yogesh Balaji, et al. Curriculum manager for source selection in multi-source domain adaptation. In *ECCV*, 2020.

[Yao *et al.*, 2021a] Xingxu Yao, Sicheng Zhao, Pengfei Xu, and Jufeng Yang. Multi-source domain adaptation for object detection. In *ICCV*, 2021.

[Yao *et al.*, 2021b] Yuan Yao, Xutao Li, et al. Multisource heterogeneous domain adaptation with conditional weighting adversarial network. *IEEE TNNLS*, 2021.

[Yin *et al.*, 2022] Yueming Yin, Zhen Yang, Haifeng Hu, and Xiaofu Wu. Universal multi-source domain adaptation for image classification. *PR*, 2022.

[Yuan *et al.*, 2022] Jin Yuan, Feng Hou, et al. Self-supervised graph neural network for multi-source domain adaptation. In *ACM MM*, 2022.

[Zhang *et al.*, 2022] Dan Zhang, Mao Ye, Yiguang Liu, Lin Xiong, and Lihua Zhou. Multi-source unsupervised domain adaptation for object detection. *INFFUS*, 2022.

[Zhang *et al.*, 2023] Wen Zhang, Lingfei Deng, Lei Zhang, and Dongrui Wu. A survey on negative transfer. *IEEE/CAA JAS*, 2023.

[Zhao *et al.*, 2018] Han Zhao, Shanghang Zhang, et al. Adversarial multiple source domain adaptation. In *NIPS*, 2018.

[Zhao *et al.*, 2019] Sicheng Zhao, Bo Li, et al. Multi-source domain adaptation for semantic segmentation. In *NeurIPS*, 2019.

[Zhao *et al.*, 2020] Sicheng Zhao, Guangzhi Wang, et al. Multi-source distilling domain adaptation. In *AAAI*, 2020.

[Zhao *et al.*, 2021a] Sicheng Zhao, Bo Li, et al. Madan: multi-source adversarial domain aggregation network for domain adaptation. *IJCV*, 2021.

[Zhao *et al.*, 2021b] Sicheng Zhao, Yang Xiao, et al. Curriculum cyclegan for textual sentiment domain adaptation with multiple sources. In *WWW*, 2021.

[Zhao *et al.*, 2022] Sicheng Zhao, Xiangyu Yue, et al. A review of single-source deep unsupervised visual domain adaptation. *IEEE TNNLS*, 2022.

[Zhao *et al.*, 2023] Sicheng Zhao, Xiaopeng Hong, et al. Toward label-efficient emotion and sentiment analysis. *Proc. IEEE*, 2023.

[Zhao *et al.*, 2024] Sicheng Zhao, Huizai Yao, et al. Multi-source-free domain adaptive object detection. *IJCV*, 2024.

[Zhou *et al.*, 2022] Lihua Zhou, Mao Ye, et al. Prototype-based multisource domain adaptation. *IEEE TNNLS*, 2022.

[Zhu *et al.*, 2019] Yongchun Zhu, Fuzhen Zhuang, and Deqing Wang. Aligning domain-specific distribution and classifier for cross-domain classification from multiple sources. In *AAAI*, 2019.

[Zuo *et al.*, 2021] Yukun Zuo, Hantao Yao, and Changsheng Xu. Attention-based multi-source domain adaptation. *IEEE TIP*, 2021.