Open World Domain Adaptation for Computer Vision

Tarun Kalluri
CSE Department, UC San Diego
sskallur@eng.ucsd.edu

Abstract

While deep learning algorithms have seen several recent successes owing to their superiority in extracting rich knowledge and structure from large scale data, most existing algorithms work with a restrictive assumption that train and test data follow similar distributions. However, distribution shift is a common phenomenon in the real world, and generalization to out-of-distribution samples at test time is still a challenging task. To address this limitation, unsupervised domain adaptation methods have been developed that utilize only unlabeled data from target domains to bridge the domain gap, with a goal to learn domain-agnostic, yet discriminative, visual representations.

In this report, I present an overview of recent literature in open world domain adaptation methods developed for computer vision. Specifically, I will discuss recent advances that go beyond conventional closed-world assumptions in domain adaptation, by relaxing the constraints on label spaces of source and target domains as well as availability of source and target data. Specifically, I will cover relaxations on (i) purely covariate shift between domains, (ii) known and fixed label spaces across domains, (iii) access to source data during adaptation, and (iv) stationary target domain. I will also talk about limitations of existing works and possible future avenues for research exploration.

1. Introduction

The role of deep learning algorithms as a driving catalyst in the field of machine learning in the last few years is unprecedented. Various applications in *computer vision* (image recognition [49, 61, 135, 141], scene understanding [21, 37, 48, 115], video analytics [133, 144, 159], object/person tracking [8, 111, 126, 150]), *natural language processing* (machine translation [67, 157, 176], textual question answering [62, 156], text summarization [25, 58, 99, 129], chatbots [35]), *reinforcement learning* (robotics [94, 98], playing games [95]) and *biotechnology* (protein folding [128], healthcare [101]) have benefited greatly due to advances in deep learning. The availability of large scale data like text [93] and images [26, 28, 78], combined with advanced training recipes [54, 136], neural network architectures [49, 149] and development of custom hardware for parallel computation have played a crucial part in driving this positive development.

A major success of these methods is due to a learning paradigm called *supervised learning*, in which a learner is provided with both the training *data* as well as the corresponding *labels* (the correct ground truth corresponding to each data point). For example, in the task of image classification, this might in the form of an (image, label) pair where each image is tagged with the correct label from a set of predefined categories (like train, bear, monkey, backpack etc). A common means to obtain these labels is through manual annotation, in which human workers are shown these images from the training set and asked to put them into one of the predefined categories, optionally introducing consistency between different workers to mitigate possible human errors. However, it is a cumbersome effort to involve human annotation for all kinds of tasks and for all the scenarios. For example, the task of semantic segmentation in computer vision involves labeling each pixel in an image with a category label, and the labels for such a task would naturally require pixel-level annotations for each image. This is very time-taking and costly (90 min/10$ per image on Cityscapes [26]). Moreover, the image scenes annotated from a particular geography cannot be *reused* to train models on a different geography, necessitating separate annotation efforts for each distinct domain of interest. Another example can be found in the task of face recognition or verification commonly found in surveillance engines, where the majority of existing labeled datasets [53, 59, 81] consist of faces from very few ethnic/racial profiles which is only a tiny subset compared to real world diversity. The models trained on these datasets alone do not deliver superior performance across demographies of people, and it is not feasible to collect representative samples from all possible ethnicities. Therefore, it is imperative to explore the possibility of using only unlabeled data which are rather easier to collect while leveraging labeled data from a different domain or distribution. However, machine learning models trained

Figure 1. Few example settings of applications of domain adaptation. In the first setting, a face recognition model which is generally trained on datasets with a majority of Caucasian faces need to be transferred to work on faces of other ethnic origin [177]. In the second case, an autonomous driving datasets may be trained on easily acquired synthetic data, but needs to be adapted before deployment on real scenes consisting of various types of climate and geographical conditions [51]. Domain adaptation provides useful techniques to improve performance on unlabeled target dataset by leveraging labels from a related source domain.

on one domain or distribution do not transfer well to samples from out of distribution, due to a property called dataset bias [143].

Domain adaptation emerged as an attractive tool to tackle this issue, where a model can be trained on datasets with little or no labels, while leveraging labels from another related domain with abundant labeled data [3, 4, 12, 13, 34, 38, 51, 55, 83–87, 91, 92, 103, 123, 131, 139, 146, 147, 175]. As a standard, the labeled domain is usually called the *source domain*, while the unlabeled domain is called the *target domain*, and the goal in domain adaptation is to minimize the error or *risk* (or equivalently improve accuracy) in the target domain. For instance, in the example segmentation scenario described above, a synthetic dataset can be easily generated with labels consisting of various traffic scenes using game engines [116, 117] and this dataset can be used as source dataset for a target domain with driving scenes from real world. Note that it is much easier, cheaper and quicker to generate synthetic scenes and more accurate to generate corresponding pixel-level labels. Adaptation algorithms can be devised that bridge the distribution or appearance shift between synthetic and real images, to achieve decent performance on real images without labels. Similarly, face collections from a population with dominant demography may be suitable as a source dataset to train models that perform on different minority population where data might be hard to acquire. In this report, we only focus on domain adaptation studied within computer vision, although other fields within machine learning like natural language processing also enjoy a rich literature in adaptation [44].

While domain adaptation has a long history in addressing covariate shift between data streams using importance re-weighting [10, 40, 114, 132], a common thread in most recent domain adaptation algorithms is to learn domain-agnostic deep representations using labeled source domain data and unlabeled target domain data, in order to transfer a classifier that acts on the source features to the target data [13, 34, 38, 83–87, 123, 139, 146, 147]. Such line of methods have been highly effective in mitigating the domain shift by learning domain-agnostic, yet discriminative, representations from large scale data giving rise to generalized transferable task-specific representations. Numerous variants of modern domain adaptation approaches include divergence minimizers [83, 85–87, 138, 139, 171], adversarial learning based methods [13, 34, 46, 84, 110, 122, 123, 146, 173], generation based methods [12, 24, 51, 97, 125], pretext task based methods [18, 102, 158, 161] among others (Tab. 1). However, a crucial assumption in most of these methods is that the prediction categories, called the *label space*, are the same across the source and target datasets. While this has been a reasonable assumption in the early days of development of adaptation algorithms, this no longer holds true in present day and age, with rapidly evolving target applications and scenarios in real world that potentially leverage deep learning models. For example, we can easily envision a scenario where an adaptation model (say a self-driving car) is trained on a source-target pair with equal label spaces, but the model might encounter unseen objects during test time outside the category set (like a new traffic cone). Without additional consideration, the model might predict the unseen unknown category erroneously to one of known categories leading to potential harm. Similarly, a very large scale source model might be needed to be used as a source dataset, while the target domain only consists of a subset of the original classes. In other cases, source data might not be accessible during adaptation due to privacy concerns or proprietary ownership reasons, and only the source trained model should be used to perform a hypothesis transfer. Acknowledging these realistic scenarios, recent works push towards addressing these challenges, which we bracket under the common umbrella called *Open World Adaptation*.

Open world adaptation is an emerging field, with a myriad of sub categories such as partial adaptation, openset adaptation, universal adaptation and source-free adaptation etc. Open world adaptation is also distinct from a more classical problem in computer vision called open world recognition [5, 82], where the latter is specifically focused towards

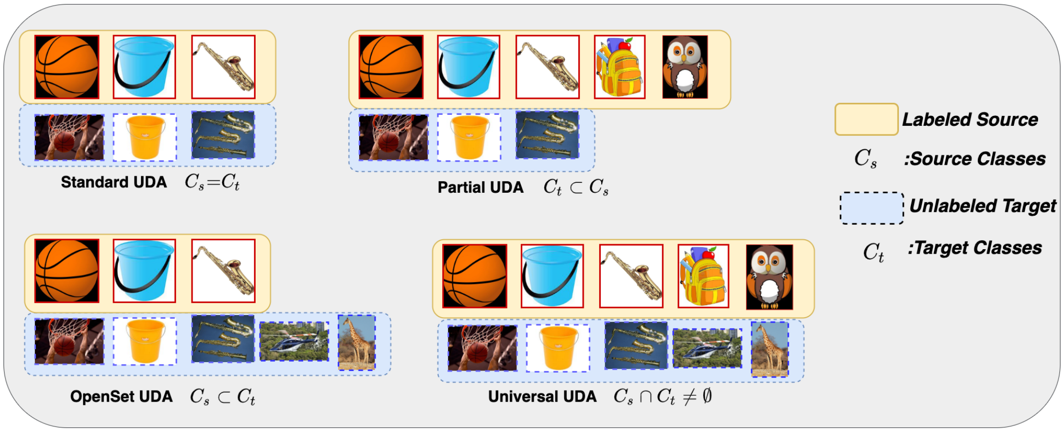

Figure 2. The various settings in domain adaptation give rise to different relationship settings. C_s corresponds to the source classes, while C_t corresponds to the target classes.

recognizing novel categories during test-time from images from the *same* domain. Open world adaptation is clearly a more challenging problem, where not only the closed world assumption of matching label spaces does not hold, but also the distribution of data at train and test time are completely different. Furthermore, open world adaptation also opens the door to solving broader problems in machine learning such as OOD detection and generalization. Fig. 2 provides a summary of relation between these various types of adaptation algorithms, defining various relaxations of the classical adaptation setting. In this report, my aim is to present a comprehensive overview of various works which are proposed in the recent past that are variants of such open world adaptation problem. These works address various relaxations of the original setting, such as (i) purely covariate shift between domains (Sec. 4.1, Sec. 4.2), (ii) known and fixed label spaces across domains (Sec. 4.3), (iii) access to source data during adaptation (Sec. 4.5), and (iv) stationary target domain (Sec. 4.6). This report is organized as follows. Sec. 2 contains the notation used in the rest of the paper. Sec. 3 contains the standard adaptation methods under the closed world setting divided into unsupervised (Sec. 3.2) and semi-supervised methods (Sec. 3.3) followed by major limitations of such methods in Sec. 3.4. Sec. 4 contains the various approaches of under the open world setting which is the theme of this report. Few limitations that still remain are discussed in Sec. 5 before concluding in Sec. 6.

2. Assumptions and Notation

In this section, the detailed notation used in the report is presented, along with the common assumptions in unsupervised domain adaptation literature. Firstly, in domain adaptation, we are given a *labeled source dataset* called D_s, containing N image-label pairs given by $X_s, Y_s = \{x_s^i, y_s^i\}_{i=1}^N$. The images are assumed to be sampled from a marginal distribution P_s. Likewise, we are also given a *target dataset* called D_t containing M images $X_t = \{x_t^i\}_{i=1}^M$ without labels, and the target samples are drawn from a distribution P_t. Optionally, the target dataset might also contain labels $Y_t = \{y_t^i\}_{i=1}^{M'}$, where M' is the number of target labels. In unsupervised domain adaptation, which is what majority of works tackle, $M' = 0$. However, there are also other settings like few-shot adaptation, semi-supervised adaptation or transfer learning where $M' > 0$ to various degrees. In any case, the goal of a learner is to use X_s, Y_s and X_t (and optionally Y_t if given) to optimize performance (or minimize error) on the unseen target test set. A running assumption in this report, unless otherwise explicitly stated, is that the marginal distributions of the source and target data are not the same, that is $P_s \nsim P_t$. However, we do assume that $P_s(y|x) \sim P_t(y|x)$, that is the conditional distributions of the labels of source and target domains are the same, allowing for meaningful knowledge transfer. For instance, in driving datasets, the source might contain images from daytime and target data can contain images from nighttime which are clearly different in their appearance and style. However, the conditional probability that a person is correctly recognized is still the same irrespective of the type of scene style. This assumption is fundamental to unsupervised domain adaptation, called as the *covariate shift assumption* [132]. Other settings include *prior shift* when $P_s(y) \nsim P_t(y)$ and concept shift when $P_s(y|x) \nsim P_t(y|x)$, but covariate shift is the most commonly studied. Also, domain adaptation is an example of *transductive transfer learning*, wherein labeled data are available only in a source domain.

Furthermore, we assume that all the source images belongs to the source category set C_s while the target images belong to the target category set C_t. While classical UDA works assume that C_s=C_t, recent open world adaptation

	Category	Assumptions	Data Requirement X_s	Y_s	X_t	Y_t	Sub-category	Methods
	Transfer Learning	$C_s \neq C_t$, $\|C\| \geq 0$	✓	✓	✓	✓	Supervised Transfer	[6, 7, 104, 130]
			✓	✗	✓	✓	Self-supervised learning	[19, 22, 36, 41, 100, 108, 109, 148, 169, 172]
Closed World Adaptation	Unsupervised Domain Adaptation (UDA)	$C_s = C_t$, $C = C_s = C_t$	✓	✓	✓	✗	Discrepancy based	[38, 83, 86, 87, 138, 139, 153] [2, 31, 56, 63, 69, 85, 96, 147, 163, 171]
							Adversarial methods	[13, 20, 34, 46, 84, 110, 122, 123, 146, 154, 155, 173] [57, 89, 106, 112, 121, 131, 145, 160]
							Generation based methods	[12, 24, 51, 97, 125]
							Self-supervised / pretext task based	[18, 102, 158, 161]
							Optimal Transport based	[9, 27]
							Other	[23, 29, 42, 68, 74, 113, 127, 134, 162]
	Semi-supervised domain adaptation	$C_s = C_t$, $C = C_s = C_t$	✓	✓	✓	✓*	Entropy Based	[55, 118, 151]
	Few-shot adaptation	$C_s = C_t$, $C = C_s = C_t$	✓	✓	✓	✓*	Fewshot adaptation	[168, 174]
Open World Adaptation	Unsupervised Domain Adaptation (UDA)	$C = C_t \subset C_s$	✓	✓	✓	✗	Partial Adaptation	[14, 16, 17, 52, 60, 72, 76, 164, 170, 178]
		$C = C_s \subset C_t$	✓	✓	✓	✗	Openset Adaptation	[1, 15, 32, 79, 88, 105, 107, 124]
		$C_s \neq C_t$, $\|C\| > 0$	✓	✓	✓	✗	Universal Adaptation	[33, 55, 70, 77, 119, 120, 167]
		$C_s \neq C_t$, $\|C\| > 0$	✗	✗	✓	✗	Source-free Adaptation	[64, 65, 71, 75, 152, 165, 166]
		$C = C_s \subset C_t$	✗	✓*	✓	✓*	Class-Incremental Adaptation	[11, 66, 80]
	Few Shot adaptation	$C_s \neq C_t$, $\|C\| = 0$	✓	✓	✓	✓*	Cross Task Transfer	[90]

Table 1. Representative works of works in domain adaptation used to learn models capable of working across domains. In all cases, the distribution of source sampled P_s is not equal to the distribution of the target samples P_t. C_s denotes the source classes and C_t denotes the target classes and C denotes the common classes between source and target. ✓ and ✗ indicate availability and non-availability of data in that particular setting respectively. ✓* indicates partial availability of labels, for example in few-shot settings.

methods relax this assumption to cover various dynamics of relationship between source and target label spaces. In this regard, we define the *common classes* between source and target using $C = C_s \cap C_t$, the source-private classes using $\overline{C_s} = C_s \setminus C$ and the target-private classes using $\overline{C_t} = C_t \setminus C$. In open world setting, the covariate shift assumption still holds, but only over the marginal distribution on common label space C.

$$P_s(x|y(x) \in C) \neq P_t(x|y(x) \in C)$$

The separation of kinds of transfer arising from different relationship between the source and datasets, along with corresponding works in literature, is presented in Tab. 1.

3. Background: closed world adaptation

As shown in Tab. 1, the closed world adaptation setting corresponds to the case when the source and the target distribution have exactly the same label spaces. These works can be further split into a number of sub-categories dependent upon the nature of the algorithm used for adaptation, namely discrepancy minimization based methods, adversarial methods, generative models, self-supervised adaptation etc. In this section, we present a short introduction to each of these works, before moving onto the main theme of the report, which is Open World adaptation, in the next section (Sec. 4). We further split the closed world adaptation works into two subsections, namely unsupervised adaptation and few-shot adaptation works depending on whether $M' = 0$ or $M' > 0$ respectively (M' is the amount of labeled target data available.)

3.1. Theoretical background

The current paradigm of unsupervised domain adaptation (hereafter UDA) is established in the pioneering work by Ben-David et. al. [3], and we present a summary of the related findings in that work here. The work first defined the $\mathcal{H}$-divergence $d_{\mathcal{H}}(D_s, D_t)$ between the source and target datasets as follows.

$$d_{\mathcal{H}}(D_s, D_t) = 2 \sup_{h \in \mathcal{H}} |\Pr_{D_s}[I(h)] - \Pr_{D_t}[I(h)]| \quad (1)$$

In simple words, the $\mathcal{H}$ divergence measure gives the maximum possible disagreement of a hypothesis on the two domains. Intuitively, if the domains are alike, then most hypotheses should do equally well (or equally bad) on both domains, so the value in above equation will be small. An attractive feature about this divergence measure is that it can be computed completely from unlabeled data, which suits the unsupervised adaptation methods. With this definition, the authors provide a target error bound as follows.

Theorem 1. *Let $\mathcal{H}$ be a hypothesis space of VC-dimensions d. If X_s and X_t are unlabeled samples of size m' each, drawn from D_s and D_t respectively, then for any $\delta \in (0, 1)$, with probability at least $1 - \delta$, for every $h \in \mathcal{H}$,*

$$\epsilon_T(h) \leq \epsilon_S(h) + \frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(D_s, D_t) + 4\sqrt{\frac{2d\log(2m') + \log(\frac{2}{\delta})}{m'}} + \lambda \quad (2)$$

The complete proof is given in [3]. Here, $\epsilon_S(h)$ and $\epsilon_T(h)$ denote the *empirical* error of source and target respectively, and λ denotes the error of the best possible joint hypothesis on source and target from the hypothesis class. Since this is an agnostic PAC setting, we assume that there exists a model that achieves this bound. Note that we also assume that $\epsilon_S(h)$ can be made very small due to the availability of source data (except in source-free settings Sec. 4.5), so the goal of most works is to reduce the divergence measure $d_{\mathcal{H}}$ between the data points to induce knowledge transfer. We now introduce various approaches for UDA that aim to minimize the target error as given in (Eq. 2).

3.2. Unsupervised adaptation methods

3.2.1 Discrepancy based methods

A large number of works aim to reduce some notion of discrepancy between the domains, and a popular choice for the discrepancy measure is the Maximum Mean Discrepency (MMD) [39]. Given a probability measure P, and a positive definite real valued kernel K, the mean embedding of the distribution is given by

$$\mu_P := \int k(., x) dP(x), \quad (3)$$

assuming a *finite-energy* kernel, that is $\int_{\mathcal{X}} k(x,x)dP(x) < \infty$. With this definition of a mean embedding, the maximum mean discrepency is defined as

$$\text{MMD}_k(P, Q) = \|\mu_P - \mu_Q\|_{\mathcal{H}}. \quad (4)$$

In practice, minimizing the MMD involves computing a deep feature space mapping for images from both the source and target domains, and then reduce the kernel mean shift between the datasets empirically computed from the samples [142]. These line of adaptation works can be grouped as MMD-based adaptation methods [56, 83, 85–87, 96, 147, 163]. Deep adaptation networks [83] use multiple kernels to compute the empirical kernel mean embedding, and then perform adaptation using features computed from multiple levels of deep layers with the intuition that the deep features progressively grow from domain specific to domain agnostic in deeper layers. Similarly, joint adaptation networks [87] advance this idea by using a tensor product in the subspace mapping stage on the *joint distribution* of all the layers at once instead of separate adaptation for each. To further address the class imbalance in most datasets, [163] proposes a sample level re-weighting mechanism during MMD adaptation. While these works are mostly aimed at aligning global domain level information, CAN [56] first uses a reliable clustering mechanism to identify pseudo-labels for target samples, and then performs *class-wise* MMD adaptation across source and target for each class separately, thereby alleviating the issue of negative transfer across classes.

While MMD is the most popular choice, it is by no means the only divergence measure used. Indeed, multiple other divergence measures have been proposed, which rely on the the *covariance discrepency*, which, in addition to the first order mean discrepency, also minimizes the second order (or optionally higher order) covariance measures between the samples from the data [2,31,96,137–139]. This is empirically found to give a better guidance to challenging adaptation tasks. Furthermore, other kinds of statistical measures have also been explored in the literature such as geodesic paths [38] or sliced Wasserstein distance metrics [69] between the datasets among others [38,63,69,153,171].

3.2.2 Adversarial learning methods

The core idea behind adversarial approaches stem from the domain divergence theory proposed in [3] and discussed briefly in Sec. 3.1. While the $\mathcal{H}-$divergence measure (Eq. 1) relies on the capacity of the hypothesis class $\mathcal{H}$ to distinguish between source and target examples, it can also be shown that one can compute the *empirical* $\mathcal{H}-$*divergence* between two finite samples $X_s \sim D_s$ and $X_t \sim D_t$ by computing

$$d_{\mathcal{H}}(X_s, X_t) = 2\left(1 - \min_{h\in\mathcal{H}}\left[\frac{1}{N}\sum_{i\in X_s} I[h(x_i) = 0] + \frac{1}{M}\sum_{j\in X_t} I[h(x_j) = 1]\right]\right), \quad (5)$$

where $I[a]$ is the indicator function which is 1 if the predicate is true and 0 otherwise. [3] also proved a uniform convergence property for the empirical estimate. When using a deep neural network, the adaptation is usually performed on a reduced dimensional deep feature space instead of the original dataspace. As such, let $S(G_f)$, $T(G_f)$ represent the computed internal representations of the neural network encoder $G_f : \mathbb{R}^{H,W,3} \rightarrow \mathbb{R}^d$. Then based on (Eq. 5), the empirical $\mathcal{H}$-divergence of a symmetrical hypothesis class is given by

$$d_{\mathcal{H}}(S(G_f), T(G_f)) = 2\left(1 - \min_{h\in\mathcal{H}}\left[\frac{1}{N}\sum_{i\in X_s} I[h(G_f(x_i)) = 0] + \frac{1}{M}\sum_{j\in X_t} I[h(G_f(x_j)) = 1]\right]\right), \quad (6)$$

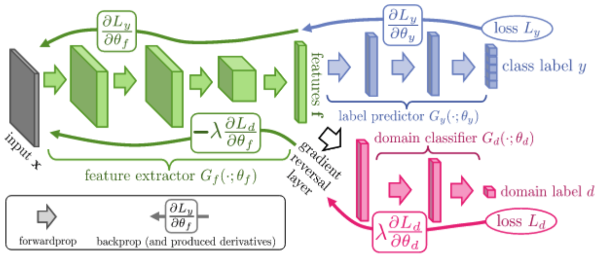

Figure 3. **DANN framework [34]** A Gradient Reversal Layer (GRL) is used to train a domain discriminator and feature encoder in an adversarial fashion. The network learns domain invariant, yet discriminative features, improving the adaptation on target data. See Sec. 3.2.2 for details.

where we replace the hypothesis on images with hypothesis acting on the feature space. The key novelty in most of the adversarial approaches is to realize the `min` term inside (Eq. 6) using a domain classifier (also called a domain discriminator) which is trained to distinguish between the domains. Simultaneously, the encoder is trained to reduce the divergence between source and target data in the feature space, so it is trained to minimize the complete term in (Eq. 6) (or maximize the term inside the bracket), giving rise to the following min-max regularization objective.

$$\begin{aligned}(\hat{\theta}_f, \hat{\theta}_y) &= \arg\min_{\theta_f, \theta_y} E(\theta_f, \theta_y, \hat{\theta}_d) \\ \hat{\theta}_d &= \arg\max_{\theta_d} E(\hat{\theta}_f, \hat{\theta}_y, \theta_d),\end{aligned} \tag{7}$$

where (θ_f, θ_y) are the parameters of the encoder model, while θ_d are the parameters of the domain discriminator. The pipeline of one such method DANN [34] is presented in Fig. 3. During training, a domain discriminator G_d is trained to reliably distinguish between the two domains when presented with the feature vectors G_f, while the encoder fools the discriminator into assigning wrong labels for the source and target samples, thereby achieving feature representations that are indistinguishable (by a good discriminator) between domains. Additionally, the feature vectors are also passed through a classifier G_y to classify them into one of the predefined categories. The min-max objective is achieved with the aid of a gradient reversal layer (GRL), which consists of an identity mapping in the forward pass, and a sign switch in the backward pass.

Adversarial adaptation methods have been the most successful, and most studied methods for unsupervised domain adaptation, as shown in Tab. 1. ADDA [146] replaces the GRL layer with a minmax GAN objective such that the training is now performed as a two player game between the feature encoder and the discriminator. CDAN [84] is a follow up work which passes the class conditional distribution for an image, in addition to the features, as input to the discriminator. Similarly, DSN [13] disentangle the private properties of each domain from the shared properties between the domains in the feature space by combining a reconstruction objective along with an orthogonality constraint to adapt only the shared feature space. Other methods propose further constraints on the alignment objective using domain discriminators [46, 112, 122].

While aforementioned methods aim at *global alignment objective*, it is also found beneficial to perform category specific alignment between the source and target domains. However, the challenge is that the target samples are unlabeled, so the class information is unknown. Therefore, methods like [106, 160] and [121] first predict psuedo labels and then align the class conditional distributions across source and target. [110] uses per-class discriminator to achieve finer level alignment, at the cost of increased network complexity. More recent works like [131, 155] perform a finer level of alignment using sample similarity criterion, along with a metric learning objective that is especially suited for datasets with finer-grained classes.

3.2.3 Generative models

The two broad varieties of methods described in Sec. 3.2.1 and Sec. 3.2.2 are called *discriminative* adaptation methods, because the alignment is performed in conjunction with training a domain agnostic classifier. However, some work follow a more *generative* adaptation approach, where the target images are first transformed to look like source images so that a classifier trained on source might be applicable to the target images [12]. However this method increases the inference time, as it additionally involves an image translation step in between. So to alleviate this drawback, works like [125] first translate source image to look like target domains, and then use the corresponding source labels to train a target specific classifier. Combining both these ideas, cycle consistency has been explored as an effective technique towards generative adaptation [24, 51, 97]. In general, image translation or generative models are found to be lesser effective and give fewer returns for the additional training complexity they introduce, while also being heavily dependent on high resolution images for effective image translation.

3.2.4 Other methods

While most adaptation approaches follow on of the aforementioned approaches, they are by no means exhaustive. Normalizing the feature norm of source and target data is found to improve adaptation in [162]. Optimal transport optimization methods have been used in [9, 27]. A spherical clustering loss is employed instead of adversarial loss in [42], while the issue of entropy regularization and consistency has also been explored [113]. Self-supervised method have also been used to good effect in [18, 140, 158, 161].

3.3. Semi-supervised methods

In semi-supervised methods, we assume that we have some amount of target labeled data present, which provides a rich guidance towards feature alignment. Note that the amount of target supervision is assumed to be only just enough to extract class properties but not enough to train a dedicated target-only classification model (which would render adaptation unnecessary).

A min-max entropy regularization objective is presented in [118, 151], where the adversarial objective is acted upon an entropy measure rather than a domain discriminator. A semi-supervised, universal adaptation model is presented in [55] to learn unifying representations across widely varying domains. Episodic training and a curriculum strategy is proposed in [174].

3.4. Limitations of closed world assumption

Although closed world adaptation models, where the label spaces of source and target are the same, have been effective at mitigating domain shift on most datasets, they do not accurately reflect the real world scenarios where the relationship between the category sets might be very different from source to target transfer. A more suitable setting is open world adaptation which is slowly gaining all the attention from the community, which we focus on next.

4. Open world adaptation, $C_s \neq C_t$

Consider a scenario where self-driving car is trained using driving data collected from US roads, like cityscapes [26]. While this model performs well on roads in the US owing to the similarity in training distribution, it does not quite work on roads in another geography, like US or India, owing to the distribution difference. Even if we did train our model to adapt to scenes from the Europe using closed world assumption (Sec. 3), it is still not sufficient to reliably deploy this model. This is because European roads contain many objects (like types of traffic cones or barricades) which might not appear on US roads, and therefore not labeled. Alternatively, a model might be trained on synthetic data which is created to contain all possible categories which can ever occur in a driving scene worldwide, but only a subset of these might actually appear in any particular target domain, in which case adaptation only needs to be performed on the smaller subset. These relaxations offer a wider opportunity to employ domain adaptation algorithms in practical real world settings. We present a suite of such recent methods under Open World Domain adaptation category. Note that the aforementioned **global alignment approaches cannot work in this setting, and in fact give rise to negative alignment**, necessitating separate treatment of this setting[1]. The possible relation between source and target categories in various settings is presented in Fig. 2.

4.1. Partial domain adaptation

Firstly, we consider the case of partial domain adaptation where the source domain contains much higher number of categories than might be potentially present in the target domain. Therefore, $C_t \subset C_s$.

Importance of the problem. In recent times, there has been a push towards creating large scale datasets with thousands of categories [28, 45] which can be reused as a source domain for a variety of target applications, without necessitating the requirement for custom source domains for each target domain. For example, a source model trained on Imagenet (which, among other classes, also contains animals and household objects) can either be used as a source data for adaptation to a target datasets which contains only animals, as well as to a target dataset which contains only household objects.

Major Challenges. Note that if we did know what categories the target might contain, then we can simply sample the source data pertaining to these classes and reduce the problem to standard adaptation procedure. However, the major constraint is the target can potentially contain any subset of source classes which is unknown at train time. Therefore, it is important to *automatically reweight* the source classes so that only those source classes which are identified in the target can participate in the adaptation procedure.

Approaches. One of the pioneering work to realize this idea is PADA [16], which computes a per-class weighting of source categories, and uses this weights to adapt only those classes that overlap with target. From (Eq. 7) we can rewrite the overall training objective for adversarial adaptation, including a re-weighting mechanism as:

$$\begin{aligned} C(\theta_f, \theta_y, \theta_d) = \frac{1}{N} \sum_{x_i \in D_s} \gamma_{y_i} L_y(G_y(G_f(x_i)), y_i) \\ - \Bigg[\frac{1}{N} \sum_{x_i \in D_s} \gamma_{y_i} L_d(G_d(G_f(x_i)), d_i) \\ + \frac{1}{M} \sum_{x_i \in D_t} L_d(G_d(G_f(x_i)), d_i) \Bigg] \end{aligned} \tag{8}$$

where the L_y is the standard classification loss over all the source samples (which also includes external categories),

[1] Negative transfer is when a private class in target domain is aligned with a wrong class in source domain, instead of forming a separate cluster, reducing the accuracy of predictions

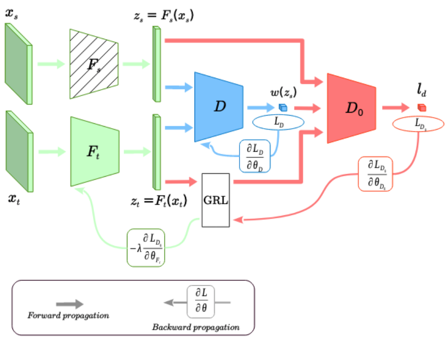

Figure 4. **IWAN** The framework of Importance weighting for partial domain adaptation, where dual-domain discriminators are used to computer domain alignment and domain similarity respectively. See Sec. 4.1 for details. Picture taken from [170].

and the term inside bracket is the domain discriminator loss, or the adversarial loss. The domain label d_i is 0 for target samples and 1 for source samples. A novel addition is the weighting parameter γ_{y_i} which indicates how likely is the corresponding label y_i to appear in the target domain. For overlapping classes ($y_i \in C$), $\gamma_{y_i} \to 1$, and for source private classes ($y_i \in \overline{C_s}$), $\gamma_{y_i} \to 0$, so that the source samples do not participate in the loss. However, target data is unlabeled, so it is not straightforward to identify the overlapping classes.

Weighting parameter. The major intuition behind designing the weighting parameter γ_{y_i} is that the target samples have higher activations on classes which overlap with the source. Therefore, a model is first trained to correctly classify the labeled source data D_s. Then, all the target data X_t is passed through the model, and the softmax activation are collected and averaged. Hence, the weight indicating the contribution of each source class to the training can be calculated as follows

$$\gamma = \frac{1}{M}\sum_{i=1}^{M}\sigma(G_y(G_f(x_i))), \tag{9}$$

where $\sigma(.)$ is the softmax activation function and γ is a C_s dimensional vector quantifying the contribution of each source class. This way, we can be ensured that the elements of γ corresponding to C have higher value than the elements of γ corresponding to $\overline{C_s}$.

However, softmax predictions can be notoriously overconfident even on OOD samples [43], so a follow-up of this idea explored using *domain similarity* as a cue for adaptation [170]. As shown in Fig. 4, they train not one, but *two* domain discriminators - both of which take as input the deep features and are trained to classify between the samples. However, only one of the domain discriminators D_o is trained in adversarial fashion to align source and target domains. The other domain discriminator (D in Fig. 4) is trained independently from rest of the network (i.e, the gradients are not backpropogated to the feature encoder). This way, the output of the second discriminator D is used to indicate the sample specific weight w_i. Specifically, the non-adversarial discriminator will output confident predictions for samples from source-private classes, while it will be less confident on overlapping classes as there is another discriminator which is acting to align these samples. Therefore, the term γ_{y_i} in (Eq. 8) is replaced by

$$w_i = 1 - D(G_f(x_i)), \tag{10}$$

while the rest of the training procedure is kept the same. This prevents pollution of source private classes into the adaptation procedure and improves the overall training accuracy. The usage of hard-weights (0/1) instead of soft weights w_i is explored in [52]. A subset selection scheme is proposed in [178], cycle consistency is used to improve alignment in [60] and Graph CNNs are used in [164]. We now look at a complementary problem of Openset adaptation.

4.2. Openset adaptation

Importance of the problem. Consider a scenario where we train an adaptation model using standard settings Sec. 3. When deployed in real world, we often cannot assume that the objects that would be encountered by the model are all seen during training. A model trained violating this fact predicts novel objects into one of pre-defined categories which might prove dangerous. Moreover, curating target data to only contain categories overlapping with source breaks the assumption of having unlabeled target data. Therefore, the models must be equipped to detect and flag novel objects as unknown during training as well as testing time. To achieve this, open set adaptation methods have been proposed.

Major Challenges. In openset adaptation, in a completely unsupervised way, we need to identify overlapping classes between source and target, and push target private classes away from common classes in feature space to avoid *negative transfer*. Simultaneously, we also need to align source and target along the common categories for effective knowledge transfer. This presents a two-fold challenge unique to this setting.

Approaches. Since the requirement of this setting is to classify known samples into one of the C categories, and flag unknown samples, the classification is performed over $C{+}1$ classes. OPDA [124] proposes to make a psuedo decision boundary to separate known and unknown classes by weakly training a separate classifier to identify all target samples as unknown class with a confidence t.

$$p(y = C_s + 1|x_t) = t \qquad 0 < t < 1 \tag{11}$$

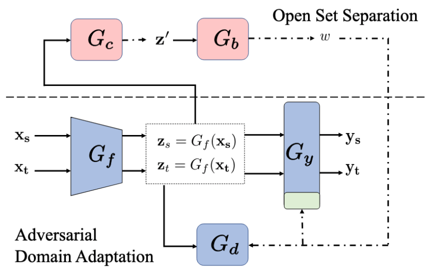

Figure 5. **Separate to Adapt** The framework of alignment architecture used in Separate to Adapt. The training proceeds in two phases, where in the first stage, the target samples are progressively separated into belonging to known and unknown classes, while in the second stage, a classifier is trained to predict target classes into either one of known or unknown category. See Sec. 4.2 for details. Picture taken from [79].

Now, the feature encoder is trained in an adversarial fashion to confuse this classifier. To effectively converge amidst such a learning criterion, they hypothesize that target samples belonging to the overlapping classes will be pushed to align with the seen categories, while target unknown samples will be pushed away from the boundary. In summary, the feature encoder will be able to choose whether a target sample should be aligned with the source or should be rejected. The adversarial loss in this case is given by

$$L_{adv}(x_t) = -t\log(p(y = C_s + 1|x_t) \\ - (1-t)\log(1 - p(y = C_s + 1|x_t)), \quad (12)$$

which is used instead of a domain discriminator. Notice that (Eq. 12) is of the form $tlogx + (1-t)\log(1-x)$, which is minimized if $x = t$. Therefore, to confuse such a classifier, all the generator needs to do is push the value of $p(y = C_s + 1|x_t)$ away from t (below t for seen classes and above t for unseen classes), which achieves the desired objective. Although this works decently, this again suffers from the problem that the predictions can be erroneously overconfident, which hurts the learning objective.

Two stage approaches. To address this issue, [79] propose a two-stage coarse-to-fine filtering mechanism to completely eliminate the influence of target unknown classes during adaptation. In the first (finer) stage, they train C_s independent binary classifiers on source data using a binary cross entropy loss to indicate if sample belongs to the corresponding classifier. Then, with the intuition that the data of known classes in the target domain tend to have higher probability in one of the shared classes than data of unknown classes, the weight for each target sample is computed as the highest probability across all of the C_s binary classifiers.

$$w_i = \max_{c \in C_s} G_c(G_f(x_i)) \qquad x_i \in D_t. \quad (13)$$

Then, in second stage (coarser) all target samples are ranked in order of highest to lowest similarity with the source domain, and the highest $n\%$ and lowest $n\%$ are chosen to train the classifier into predicting samples into one of $C+1$ classes including the unknown class. The overall pipeline is presented in Fig. 5. To calibrate the difficulty of the adaptation setting, an *openness* score is proposed which is given by

$$\mathrm{O} = 1 - \frac{|C_s|}{|C_t|}, \quad (14)$$

which lies between 0-1 indicating the range from high label overlap to very little overlap (low openness to high openness, and hence increasing difficulty).

The possibility of only using a self-supervised pretext task such as rotation prediction has been explored in [15]. Following the success of rotation prediction in anomaly detection [50]. Using a similar two stage procedure, the target samples are first segregated into two sets depending on their similarity to the source domain, computed from the rotation prediction task. In the second stage, the most confident of the predictions are used to perform alignment with source domain, or classify as unknown depending upon the weight computed from the first stage. Other approaches to tackle openset adaptation include performing selective clustering on target sample [105] and graph learning [88].

Although both partial and openset domain adaptation serve useful relaxations of the standard adaptation procedure, they still assume that the user *knows beforehand* the relationship between the labels. A more general setting is to work with completely unknown label relationship between source and target - which is what universal domain adaptation is proposed to solve.

4.3. Universal domain adaptation

Importance of the problem. Universal domain adaptation, first proposed in [167] in a broad sense combines all the settings of adaptation scenarios we have so far into a single adaptation setting. In this case, we assume that the source and target label spaces are different with finite overlap between them, but also both source and target contain a *separate private set* with categories that do not belong to the common categories. Therefore, now the model is required to (i) classify the target sample correctly if it is associated with a label in the common label set C, or (ii) mark it as 'unknown' otherwise. Similar to the openness score in (Eq. 14), the difficulty measure in this setting is calibrated using a *commonness* criterion, given by

$$\xi = \frac{|C_s \cap C_t|}{|C_s \cup C_t|} = \frac{C}{|C_s \cup C_t|} \quad (15)$$

Closed set adaptation is a special case of the universal setting when $\xi = 1$.

Major Challenges. The smaller the value of ξ, the less sharing knowledge is and more difficult the adaptation is. More importantly, universal methods needs to be applicable over a wide range of ξ, thereby encompassing all of the previous settings.

Approaches. Since the problem is a combination of partial and openset settings, [167] propose an algorithm which is also a combination of the previous algorithms. Specifically, they extend the weighting parameter in [170], given in (Eq. 10) to account for sample selection across both source and target. With $\hat{d}$ as the domain discriminator prediction, they make the following crucial observation.

$$\mathbb{E}_{x\sim\overline{C_s}}(\hat{d}) > \mathbb{E}_{x\sim C_s}(\hat{d}) > \mathbb{E}_{x\sim C_t}(\hat{d}) > \mathbb{E}_{x\sim\overline{C_t}}(\hat{d}). \quad (16)$$

This is because the samples from the source private set $\overline{C_s}$ have maximum confidence into being predicted as belonging to source. Next, since there is non-zero overlap between C_s and C_t, the samples belonging to the common classes will have slightly more confusion. Finally, the samples from target private set $\overline{C_t}$ have least score since they are confidently predicted as target. In addition to this, they also use the entropy of the prediction output (H) as a measure of the confidence, so the final weight is given by a combination of both of these

$$\begin{aligned} w^s(x) &= \frac{H(G_y(G_f(x)))}{\log(|C_s|)} - D(G_f(x)) \\ w^t(x) &= 1 - w^s(x). \end{aligned} \quad (17)$$

With this, the loss equation is now defined similar to (Eq. 8), with weights to eliminate potential outside classes from both source and target.

$$\begin{aligned} C(\theta_f,\theta_y,\theta_d) = &\frac{1}{N}\sum_{x_i\in D_s} w_s(x_i)L_y(G_y(G_f(x_i)),y_i) \\ &- \left[\frac{1}{N}\sum_{x_i\in D_s} w_s(x_i)L_d(G_d(G_f(x_i)),d_i)\right. \\ &\left.+\frac{1}{M}\sum_{x_i\in D_t} w_t(x_i)L_d(G_d(G_f(x_i)),d_i)\right] \end{aligned} \quad (18)$$

During inference, if the weight is below a validated threshold, then the sample is declared to be unknown, else it is predicted into one of the C categories. [33] proposes additional heuristics like consistency loss and classifier confidence to further strengthen the weight parameter.

In a completely different treatment to the problem, a recent approach called OVA-Net [120] proposes to use a one-vs-all classifier to separate common categories with private categories during training. They achieve this by training C_s independent binary classifier, where each classifier makes a choice if the sample belongs to its class or not, as shown in Fig. 6. Having this, during test-time, if none of the classifier votes for a sample, then the sample is deemed to be from outside any of the known categories and judged to be 'unknown'. Additionally, they also make use of a hard negative classifier sampling along with entropy minimization regularization to train the binary classifiers. A major advantage of this method is that it is free of choosing any threshold hyperparameters, which avoids the use of any target labeled data.

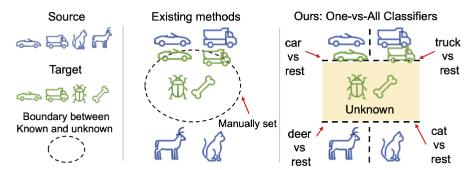

Figure 6. **OVA-Net** The framework of alignment architecture used in OVA-net. The independent binary classifiers are shown to be more robust to detecting OOD samples compared to softmax classifier. See Sec. 4.3 for details. Picture taken from [120].

Other approaches towards universal adaptation include self-supervision [119], clustering [70] and optimum sample selection [77]. The comparison table between accuracy values of all the works is presented in Tab. 2 for Office-Home and DomainNet datasets.

4.4. A note on the evaluation metrics

While traditional domain adaptation methods are generally calibrated and compared over the accuracy they deliver over the target test set, recent open world methods call for a different evaluation metric. This is because the models now not only need to be evaluated on the accuracy delivered on the common (seen) classes, but also how well they are able to reject (or abstain prediction) on unseen classes. Therefore, some works [79,124,167] just use the unknown prediction as an additional class and report accuracy over the $C+1$ classes. However, this might unfairly favor methods which have better numbers on common classes, especially in case of large scale datasets. Therefore, recent works [15, 33] use a more meaningful metric which is formed as a harmonic mean of seen class accuracy and unseen class accuracy as:

$$h = 2\frac{a_c \cdot a_{\bar{c}^t}}{a_c + a_{\bar{c}^t}}$$

where a_c is the accuracy on seen classes, while $a_{\bar{c}^t}$ is unknown prediction accuracy. This metric is high only when both the a_c and $a_{\bar{c}^t}$ are high, so that both abilities of universal adaptation methods are taken into account.

4.5. Source-free methods

Importance of the problem. In this section, we switch our focus to relax yet another assumption which is generally

Method	DomainNet (150 / 50 / 145)							VisDA (6 / 3 / 3)	OfficeHome (15 / 5 / 50)
	P2R	R2P	P2S	S2P	R2S	S2R	Avg		
DANCE [119]	21.0	47.3	37.0	27.7	46.7	21.0	33.5	4.4	49.2
UAN [167]	41.9	43.6	39.1	38.9	38.7	43.7	41.0	30.5	56.6
CMU [33]	50.8	52.2	45.1	44.8	45.6	51.0	48.3	34.6	61.6
DCC [70]	56.9	50.3	43.7	44.9	43.3	56.2	49.2	43.0	70.2
OVA-Net [120]	56.0	51.7	47.1	47.4	44.9	57.2	50.7	53.1	71.8

Table 2. **H-score (Eq. 4.4) of universal DA using DomainNet, VisDA and OfficeHome**. The numbers in brackets alongside each dataset correspond to the number of shared, source-private and target-private categories in each setting.

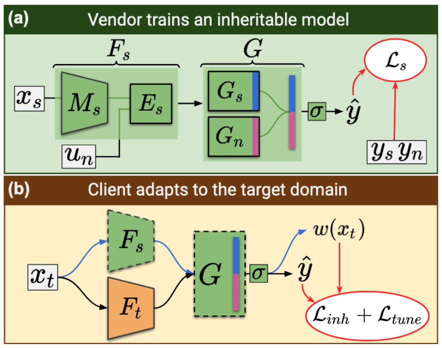

Figure 7. **Source-free adaptation training** The framework of alignment architecture proposed for source-free training in [65]. The training proceeds in two stages, wherein a foresighted source model and a source-free target classifier are trained respectively. See Sec. 4.5 for details. Picture taken from [65].

found in most adaptation methods, which is availability of source data during adaptation. This assumption of coexisting source and target datasets poses a significant constraint in the modern world [65], where coping up with strict digital privacy and copyright laws is of prime importance. For example, consider a setting where a facial recognition system is developed by a corporate house, and a user wishes to adapt the model to suit their specific demography. In this case, the vendor might be unwilling to provide the source data due to privacy issues of the individual faces, but can provide access to a model trained on source data. This gives rise to a new challenge called source-free domain adaptation setting.

Two stage approaches. The authors in [75] explore only using source hypothesis for adaptation without source data under standard adaptation settings. Recently, source-free models for openset adaptation [65] and universal adaptation [64] have been proposed. A running theme of these works is a two-stage training method. In the first stage, the vendor who has access to the source data trains a foresighted model which would hold well even when presented with adaptation to a target data. Later, the source model is used to re-weight the target samples into seen or unseen categories, and the source predictions along with a fine-tuned target classifier is used to make the final predictions.

To achieve the first step, a robust model is trained on source data along with synthetically generated OOD samples to allow room for open set target samples in the future. The synthetic OOD samples are either generated using *feature splicing* in [65], or fitting a Gaussian posterior and sampling points from outside 3σ of the Gaussian in [64]. Either way, this allows the model to be trained for posterity when presented with target domain data. A complete schematic of this method is provided in Fig. 7. Other approaches for source-free adaptation include [71, 152, 165, 166] that primarily vary in the type of weighting mechanisms used for source hypothesis transfer.

4.6. Incremental adaptation

Importance of the problem. In addition to source data availability, another related yet tricky assumption is to assume fixed or stationary target domains. In practice, targets are seldom stationary. For example, a self-driving car need to be equipped with various climatic conditions such as rainy, snowy or fogy roads at different times of the year. Adaptation to a single one of these would not suffice, neither would it be prudent to assume that all variants of the target domain would be simultaneously available.

Major Challenges. A key challenge that needs to be addressed in this setting is to prevent *catastrophic forgetting* [73], wherein a model that is trained to work well on a target domain should be able to retain its performance when presented with a new target domain, that is, it should never *forget* whatever it learnt previously.

Approaches. To deal with this problem, [11] first introduces adaptation in the presence of continuously shifting (target) domains. The paper presents a training approach in an adaptation scenario where the learner is provided with a labeled source sample and unlabeled target samples at different time intervals, as the target domain is changing over time. At every stage, the approach is to save a small subset of the predictions and replay (reuse) them as soft labels for future domains. The past soft labels are enforced to be matched using a consistency replay loss, in addition to adapting to existing target domains using standard adversarial approaches.

While the aforementioned work deals with *domain in-*

cremental learning, more recent approaches also tackle a related problem of *class incremental learning*, where the target domain remains the same but the classes in target data keep increasing. For example, some new traffic installations or road signs might be decided to be added to the roads and our model should be flexible to accommodate these changes. A related work is [66], which proposes a two-stage approach to achieve class incremental learning. In the first stage, a *foresighted* source model is trained along with negative training, while in the second stage, the target samples are clustered into known and unknown splits and adapted separately. Alternatively, a meta-learning setup is proposed in [80] to achieve adaptation amidst moving domains and a few shot learning objective is proposed in [90].

5. Limitations and possible future work

In this section, we note some of the limitations that still remain in spite of massive progress in domain adaptation in recent years. Firstly, most works that deal with open world or open set adaptation classify all the unseen classes as unknown. This is not ideal, since there might be a large variety and diversity in the unseen classes, and grouping them into a single cluster not only weakens the learning objective, but also ignores the rich structure available in the data. Moreover, if we do have annotations in the future for some of the unseen classes, then, this model would not be very accommodating and conducive to further finetuning. Therefore, to achieve better adaptation performance and incremental learning, it is important to classify seen categories into one of the known categories, while simultaneously also forming distinct (unlabeled) clusters for all the unseen objects. This is more closer to the principle of human perception, where it might be difficult for humans to recognize unknown labels for the first time, but we can easily 'group' objects into clusters based on similarity.

Secondly, current adaptation models only focus on improving and optimizing for the target accuracy, but in the process, the accuracy on the source model will be hurt due to catastrophic forgetting on source data [73], necessitating individual models for deployment in source and target domains. With the trend of ever increasing size of deep learning models [30], it becomes cumbersome to deploy multiple models on resource limited hardware preventing universal deployment of models. Therefore, the efforts should also be directed towards building models that not only perform well on target after adaptation, but also retain their performance on the source data. Again, this is very similar to how human knowledge works. We continuously aggregate new information while having a special ability to retain our relevant past knowledge.

Thirdly, there is no uniform way of calibrating the performance of models without having access to target domain data. Most of the adaptation works, although they work with the assumption of completely unlabeled target, nevertheless use target validation data for model selection which poses a risk of over-fitting to a specific target domain. Therefore, it is unclear currently how well these methods compare against each other when deployed in real world. Considerable efforts need to be directed towards building unsupervised evaluation protocols for adaptation, on the lines of evaluating the clustering quality in category discovery [47].

6. Conclusion

In this report, we presented a comprehensive study of various open world domain adaptation methods which enable model training on unlabeled data using labels from a related source data by relaxing various impractical constraints on prior adaptation methods. After a brief overview of the theoretical and practical aspects that form the background of standard adaptation methods, we presented approaches that are developed towards partial, OpenSet and universal domain adaptation that are proposed to work with various relationship between source and target categories. Furthermore, we also discussed the ideas of source-free adaptation and adaptation with moving targets that relax the constraint of availability of source and target data respectively during training. Finally, we also discussed some limitations that still remain in current adaptation literature which needs to addressed as a priority. Finally, although this report is focused on advances in computer vision, other fields like reinforcement learning or natural language processing have also greatly benefited from similar principles in domain adaptation. We hope that this report can serve as a useful primer to the broad field of adaptation and indicator for possible future avenues of research.

References

[1] Mahsa Baktashmotlagh, Masoud Faraki, Tom Drummond, and Mathieu Salzmann. Learning factorized representations for open-set domain adaptation. *arXiv preprint arXiv:1805.12277*, 2018. 4

[2] Mahsa Baktashmotlagh, Mehrtash Harandi, and Mathieu Salzmann. Distribution-matching embedding for visual domain adaptation. *The Journal of Machine Learning Research*, 17(1):3760–3789, 2016. 4, 5

[3] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79(1-2):151–175, 2010. 2, 4, 5

[4] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. *Advances in neural information processing systems*, 19:137–144, 2006. 2

[5] Abhijit Bendale and Terrance Boult. Towards open world recognition. In *Proceedings of the IEEE conference on*

computer vision and pattern recognition, pages 1893–1902, 2015. 2

[6] Yoshua Bengio. Deep learning of representations for unsupervised and transfer learning. In *Proceedings of ICML workshop on unsupervised and transfer learning*, pages 17–36. JMLR Workshop and Conference Proceedings, 2012. 4

[7] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013. 4

[8] Luca Bertinetto, Jack Valmadre, Joao F Henriques, Andrea Vedaldi, and Philip HS Torr. Fully-convolutional siamese networks for object tracking. In *European conference on computer vision*, pages 850–865. Springer, 2016. 1

[9] Bharath Bhushan Damodaran, Benjamin Kellenberger, Rémi Flamary, Devis Tuia, and Nicolas Courty. Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 447–463, 2018. 4, 6

[10] Steffen Bickel, Michael Brückner, and Tobias Scheffer. Discriminative learning under covariate shift. *Journal of Machine Learning Research*, 10(9), 2009. 2

[11] Andreea Bobu, Eric Tzeng, Judy Hoffman, and Trevor Darrell. Adapting to continuously shifting domains. In *ICLR*, 2018. 4, 11

[12] Konstantinos Bousmalis, Nathan Silberman, David Dohan, Dumitru Erhan, and Dilip Krishnan. Unsupervised pixel-level domain adaptation with generative adversarial networks. In *CVPR*, pages 95–104. IEEE Computer Society, 2017. 2, 4, 6

[13] Konstantinos Bousmalis, George Trigeorgis, Nathan Silberman, Dilip Krishnan, and Dumitru Erhan. Domain separation networks. In *Advances in neural information processing systems*, pages 343–351, 2016. 2, 4, 6

[14] Silvia Bucci, Antonio D'Innocente, and Tatiana Tommasi. Tackling partial domain adaptation with self-supervision. *ArXiv*, abs/1906.05199, 2019. 4

[15] Silvia Bucci, Mohammad Reza Loghmani, and Tatiana Tommasi. On the effectiveness of image rotation for open set domain adaptation. In *European Conference on Computer Vision*, pages 422–438. Springer, 2020. 4, 9, 10

[16] Zhangjie Cao, Lijia Ma, Mingsheng Long, and Jianmin Wang. Partial adversarial domain adaptation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 135–150, 2018. 4, 7

[17] Zhangjie Cao, Kaichao You, Mingsheng Long, Jianmin Wang, and Qiang Yang. Learning to transfer examples for partial domain adaptation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2980–2989, 2019. 4

[18] Fabio M Carlucci, Antonio D'Innocente, Silvia Bucci, Barbara Caputo, and Tatiana Tommasi. Domain generalization by solving jigsaw puzzles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2229–2238, 2019. 2, 4, 7

[19] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 9912–9924. Curran Associates, Inc., 2020. 4

[20] Chaoqi Chen, Weiping Xie, Wenbing Huang, Yu Rong, Xinghao Ding, Yue Huang, Tingyang Xu, and Junzhou Huang. Progressive feature alignment for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 627–636, 2019. 4

[21] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4):834–848, 2017. 1

[22] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. 4

[23] Xinyang Chen, Sinan Wang, Mingsheng Long, and Jianmin Wang. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In *International conference on machine learning*, pages 1081–1090. PMLR, 2019. 4

[24] Yun-Chun Chen, Yen-Yu Lin, Ming-Hsuan Yang, and Jia-Bin Huang. Crdoco: Pixel-level domain transfer with cross-domain consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1791–1800, 2019. 2, 4, 6

[25] Jianpeng Cheng and Mirella Lapata. Neural summarization by extracting sentences and words. *arXiv preprint arXiv:1603.07252*, 2016. 1

[26] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016. 1, 7

[27] Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 39(9):1853–1865, 2016. 4, 6

[28] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 1, 7

[29] Zhijie Deng, Yucen Luo, and Jun Zhu. Cluster alignment with a teacher for unsupervised domain adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9944–9953, 2019. 4

[30] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold,

Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020. 12

[31] Yuntao Du, Ruiting Zhang, Yikang Cao, Xiaowen Zhang, Zhiwen Tan, and Chongjun Wang. Unsupervised domain adaptation with exploring more statistics and discriminative information. *arXiv preprint arXiv:2003.11723*, 2020. 4, 5

[32] Zhen Fang, Jie Lu, Feng Liu, Junyu Xuan, and Guangquan Zhang. Open set domain adaptation: Theoretical bound and algorithm. *IEEE transactions on neural networks and learning systems*, 2020. 4

[33] Bo Fu, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Learning to detect open classes for universal domain adaptation. In *European Conference on Computer Vision*, pages 567–583. Springer, 2020. 4, 10, 11

[34] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015. 2, 4, 6

[35] Jianfeng Gao, Michel Galley, and Lihong Li. Neural approaches to conversational ai. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, pages 1371–1374, 2018. 1

[36] Spyros Gidaris, Praveer Singh, and Nikos Komodakis. Unsupervised representation learning by predicting image rotations. *CoRR*, abs/1803.07728, 2018. 4

[37] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587, 2014. 1

[38] Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2066–2073. IEEE, 2012. 2, 4, 5

[39] Arthur Gretton, Karsten Borgwardt, Malte Rasch, Bernhard Schölkopf, and Alex Smola. A kernel method for the two-sample-problem. In B. Schölkopf, J. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems*, volume 19. MIT Press, 2007. 5

[40] Arthur Gretton, Alex Smola, Jiayuan Huang, Marcel Schmittfull, Karsten Borgwardt, and Bernhard Schölkopf. Covariate shift by kernel mean matching. *Dataset shift in machine learning*, 3(4):5, 2009. 2

[41] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Daniel Guo, Mohammad Gheshlaghi Azar, et al. Bootstrap your own latent: A new approach to self-supervised learning. *arXiv preprint arXiv:2006.07733*, 2020. 4

[42] Xiang Gu, Jian Sun, and Zongben Xu. Spherical space domain adaptation with robust pseudo-label loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9101–9110, 2020. 4, 6

[43] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *International Conference on Machine Learning*, pages 1321–1330. PMLR, 2017. 8

[44] Han Guo, Ramakanth Pasunuru, and Mohit Bansal. Multi-source domain adaptation for text classification via distancenet-bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020. 2

[45] Agrim Gupta, Piotr Dollar, and Ross Girshick. LVIS: A dataset for large vocabulary instance segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 7

[46] Philip Haeusser, Thomas Frerix, Alexander Mordvintsev, and Daniel Cremers. Associative domain adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2765–2773, 2017. 2, 4, 6

[47] Kai Han, Andrea Vedaldi, and Andrew Zisserman. Learning to discover novel visual categories via deep transfer clustering. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8401–8409, 2019. 12

[48] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 1

[49] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv preprint arXiv:1512.03385*, 2015. 1

[50] Dan Hendrycks, Mantas Mazeika, Saurav Kadavath, and Dawn Song. Using self-supervised learning can improve model robustness and uncertainty. *Advances in Neural Information Processing Systems (NeurIPS)*, 2019. 9

[51] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning*, pages 1989–1998. PMLR, 2018. 2, 4, 6

[52] Jian Hu, Hongya Tuo, Chao Wang, Lingfeng Qiao, Haowen Zhong, Junchi Yan, Zhongliang Jing, and Henry Leung. Discriminative partial domain adversarial network. In *ECCV*, 2020. 4, 8

[53] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*, 2008. 1

[54] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015. 1

[55] Tarun Kalluri, Girish Varma, Manmohan Chandraker, and CV Jawahar. Universal semi-supervised semantic segmentation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5259–5270, 2019. 2, 4, 7

[56] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4893–4902, 2019. 4, 5

[57] Guoliang Kang, Liang Zheng, Yan Yan, and Yi Yang. Deep adversarial attention alignment for unsupervised domain

adaptation: the benefit of target expectation maximization. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 401–416, 2018. 4

[58] Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6769–6781, Online, Nov. 2020. Association for Computational Linguistics. 1

[59] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4396–4405, 2019. 1

[60] Youngeun Kim, Sungeun Hong, Seunghan Yang, Sungil Kang, Yunho Jeon, and Jiwon Kim. Associative partial domain adaptation. *ArXiv*, abs/2008.03111, 2020. 4, 8

[61] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012. 1

[62] Ankit Kumar, Ozan Irsoy, Peter Ondruska, Mohit Iyyer, James Bradbury, Ishaan Gulrajani, Victor Zhong, Romain Paulus, and Richard Socher. Ask me anything: Dynamic memory networks for natural language processing. In Maria-Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, volume 48 of *JMLR Workshop and Conference Proceedings*, pages 1378–1387. JMLR.org, 2016. 1

[63] Abhishek Kumar, Prasanna Sattigeri, Kahini Wadhawan, Leonid Karlinsky, Rogerio Feris, Bill Freeman, and Gregory Wornell. Co-regularized alignment for unsupervised domain adaptation. In *Advances in Neural Information Processing Systems*, pages 9345–9356, 2018. 4, 5

[64] Jogendra Nath Kundu, Naveen Venkat, R Venkatesh Babu, et al. Universal source-free domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4544–4553, 2020. 4, 11

[65] Jogendra Nath Kundu, Naveen Venkat, Ambareesh Revanur, R Venkatesh Babu, et al. Towards inheritable models for open-set domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12376–12385, 2020. 4, 11

[66] Jogendra Nath Kundu, Rahul Mysore Venkatesh, Naveen Venkat, Ambareesh Revanur, and R. Venkatesh Babu. Class-incremental domain adaptation. *ArXiv*, abs/2008.01389, 2020. 4, 12

[67] Guillaume Lample, Alexis Conneau, Ludovic Denoyer, and Marc'Aurelio Ranzato. Unsupervised machine translation using monolingual corpora only. *arXiv preprint arXiv:1711.00043*, 2017. 1

[68] Issam H Laradji and Reza Babanezhad. M-adda: Unsupervised domain adaptation with deep metric learning. In *Domain Adaptation for Visual Understanding*, pages 17–31. Springer, 2020. 4

[69] Chen-Yu Lee, Tanmay Batra, Mohammad Haris Baig, and Daniel Ulbricht. Sliced wasserstein discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 10285–10295, 2019. 4, 5

[70] Guangrui Li, Guoliang Kang, Yi Zhu, Yunchao Wei, and Yi Yang. Domain consensus clustering for universal domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9757–9766, 2021. 4, 10, 11

[71] Rui Li, Qianfen Jiao, Wenming Cao, Hau-San Wong, and Si Wu. Model adaptation: Unsupervised domain adaptation without source data. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9638–9647, 2020. 4, 11

[72] Weikai Li and Songcan Chen. Partial domain adaptation without domain alignment. *ArXiv*, abs/2108.12867, 2021. 4

[73] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017. 11, 12

[74] Qing Lian, Lixin Duan, Fengmao Lv, and Boqing Gong. Constructing self-motivated pyramid curriculums for cross-domain semantic segmentation: A non-adversarial approach. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6757–6766. IEEE, 2019. 4

[75] Jian Liang, D. Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *ICML*, 2020. 4, 11

[76] Jian Liang, Yunbo Wang, D. Hu, Ran He, and Jiashi Feng. A balanced and uncertainty-aware approach for partial domain adaptation. In *ECCV*, 2020. 4

[77] Omri Lifshitz and Lior Wolf. Sample selection for universal domain adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(10):8592–8600, May 2021. 4, 10

[78] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 1

[79] Hong Liu, Zhangjie Cao, Mingsheng Long, Jianmin Wang, and Qiang Yang. Separate to adapt: Open set domain adaptation via progressive separation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2927–2936, 2019. 4, 9, 10

[80] Hong Liu, Mingsheng Long, Jianmin Wang, and Yu Wang. Learning to adapt to evolving domains. In *NeurIPS*, 2020. 4, 12

[81] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, December 2015. 1

[82] Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and Stella X Yu. Large-scale long-tailed recognition in an open world. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2537–2546, 2019. 2

[83] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015. 2, 4, 5

[84] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. In *Advances in Neural Information Processing Systems*, pages 1640–1650, 2018. 2, 4, 6

[85] Mingsheng Long, Jianmin Wang, Guiguang Ding, Jiaguang Sun, and Philip S Yu. Transfer feature learning with joint distribution adaptation. In *Proceedings of the IEEE international conference on computer vision*, pages 2200–2207, 2013. 2, 4, 5

[86] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Unsupervised domain adaptation with residual transfer networks. In *Advances in neural information processing systems*, pages 136–144, 2016. 2, 4, 5

[87] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. In *International conference on machine learning*, pages 2208–2217. PMLR, 2017. 2, 4, 5

[88] Yadan Luo, Zijian Wang, Zi Huang, and Mahsa Baktashmotlagh. Progressive graph learning for open-set domain adaptation. In *International Conference on Machine Learning*, pages 6468–6478. PMLR, 2020. 4, 9

[89] Yawei Luo, Liang Zheng, Tao Guan, Junqing Yu, and Yi Yang. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2507–2516, 2019. 4

[90] Zelun Luo, Yuliang Zou, Judy Hoffman, and Li Fei-Fei. Label efficient learning of transferable representations across domains and tasks. *arXiv preprint arXiv:1712.00123*, 2017. 4, 12

[91] Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. Domain adaptation: Learning bounds and algorithms. In *COLT*, 2009. 2

[92] Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. Multiple source adaptation and the renyi divergence. In *UAI*, pages 367–374, 2009. 2

[93] Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. The natural language decathlon: Multitask learning as question answering. *arXiv preprint arXiv:1806.08730*, 2018. 1

[94] Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Timothy P. Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In *ICML*, 2016. 1

[95] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013. 1

[96] Pietro Morerio, Jacopo Cavazza, and Vittorio Murino. Minimal-entropy correlation alignment for unsupervised deep domain adaptation. *arXiv preprint arXiv:1711.10288*, 2017. 4, 5

[97] Zak Murez, Soheil Kolouri, David Kriegman, Ravi Ramamoorthi, and Kyungnam Kim. Image to image translation for domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4500–4509, 2018. 2, 4, 6

[98] Ofir Nachum, Mohammad Norouzi, Kelvin Xu, and Dale Schuurmans. Bridging the gap between value and policy based reinforcement learning. In *NIPS*, 2017. 1

[99] Ramesh Nallapati, Bowen Zhou, Cícero Nogueira dos Santos, Çaglar Gülçehre, and Bing Xiang. Abstractive text summarization using sequence-to-sequence rnns and beyond. In Yoav Goldberg and Stefan Riezler, editors, *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning, CoNLL 2016, Berlin, Germany, August 11-12, 2016*, pages 280–290. ACL, 2016. 1

[100] Mehdi Noroozi and Paolo Favaro. Unsupervised learning of visual representations by solving jigsaw puzzles. In *European conference on computer vision*, pages 69–84. Springer, 2016. 4

[101] Ohad Oren, Bernard J Gersh, and Deepak L Bhatt. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. *The Lancet Digital Health*, 2(9):e486–e488, 2020. 1

[102] Fei Pan, Inkyu Shin, Francois Rameau, Seokju Lee, and In So Kweon. Unsupervised intra-domain adaptation for semantic segmentation through self-supervision. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3764–3773, 2020. 2, 4

[103] Sinno Jialin Pan, Ivor W Tsang, James T Kwok, and Qiang Yang. Domain adaptation via transfer component analysis. *IEEE Transactions on Neural Networks*, 22(2):199–210, 2010. 2

[104] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009. 4

[105] Yingwei Pan, Ting Yao, Yehao Li, Chong-Wah Ngo, and Tao Mei. Exploring category-agnostic clusters for open-set domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13867–13875, 2020. 4, 9

[106] Yingwei Pan, Ting Yao, Yehao Li, Yu Wang, Chong-Wah Ngo, and Tao Mei. Transferrable prototypical networks for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2239–2247, 2019. 4, 6

[107] Pau Panareda Busto and Juergen Gall. Open set domain adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 754–763, 2017. 4

[108] Deepak Pathak, Ross Girshick, Piotr Dollár, Trevor Darrell, and Bharath Hariharan. Learning features by watching objects move. In *CVPR*, 2017. 4

[109] Deepak Pathak, Philipp Krähenbühl, Jeff Donahue, Trevor Darrell, and Alexei A. Efros. Context encoders: Feature learning by inpainting. In *CVPR*, 2016. 4

[110] Zhongyi Pei, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Multi-adversarial domain adaptation. *arXiv preprint arXiv:1809.02176*, 2018. 2, 4, 6

[111] Federico Perazzi, Oliver Wang, Markus Gross, and Alexander Sorkine-Hornung. Fully connected object proposals for video segmentation. In *Proceedings of the IEEE international conference on computer vision*, pages 3227–3234, 2015. 1

[112] Pedro O Pinheiro. Unsupervised domain adaptation with similarity learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8004–8013, 2018. 4, 6

[113] Viraj Prabhu, Shivam Khare, Deeksha Kartik, and Judy Hoffman. Sentry: Selective entropy optimization via committee consistency for unsupervised domain adaptation. *arXiv preprint: 2012.11460*, 2020. 4, 6

[114] Joaquin Quiñonero-Candela, Masashi Sugiyama, Anton Schwaighofer, and Neil D Lawrence. An adversarial view of covariate shift and a minimax approach. 2009. 2

[115] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28:91–99, 2015. 1

[116] Stephan R Richter, Vibhav Vineet, Stefan Roth, and Vladlen Koltun. Playing for data: Ground truth from computer games. In *European conference on computer vision*, pages 102–118. Springer, 2016. 2

[117] German Ros, Laura Sellart, Joanna Materzynska, David Vazquez, and Antonio M Lopez. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3234–3243, 2016. 2

[118] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, Trevor Darrell, and Kate Saenko. Semi-supervised domain adaptation via minimax entropy. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8050–8058, 2019. 4, 7

[119] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, and Kate Saenko. Universal domain adaptation through self-supervision. 2020. 4, 10, 11

[120] Kuniaki Saito and Kate Saenko. Ovanet: One-vs-all network for universal domain adaptation. *arXiv preprint arXiv:2104.03344*, 2021. 4, 10, 11

[121] Kuniaki Saito, Yoshitaka Ushiku, and Tatsuya Harada. Asymmetric tri-training for unsupervised domain adaptation. *arXiv preprint arXiv:1702.08400*, 2017. 4, 6

[122] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Adversarial dropout regularization. *arXiv preprint arXiv:1711.01575*, 2017. 2, 4, 6

[123] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3723–3732, 2018. 2, 4

[124] Kuniaki Saito, Shohei Yamamoto, Yoshitaka Ushiku, and Tatsuya Harada. Open set domain adaptation by backpropagation. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 153–168, 2018. 4, 8, 10

[125] Swami Sankaranarayanan, Yogesh Balaji, Carlos D Castillo, and Rama Chellappa. Generate to adapt: Aligning domains using generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8503–8512, 2018. 2, 4, 6

[126] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015. 1

[127] Ozan Sener, Hyun Oh Song, Ashutosh Saxena, and Silvio Savarese. Learning transferrable representations for unsupervised domain adaptation. In *Advances in Neural Information Processing Systems*, pages 2110–2118, 2016. 4

[128] Andrew W Senior, Richard Evans, John Jumper, James Kirkpatrick, Laurent Sifre, Tim Green, Chongli Qin, Augustin Žídek, Alexander WR Nelson, Alex Bridgland, et al. Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710, 2020. 1

[129] Min Joon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. Bidirectional attention flow for machine comprehension. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017. 1

[130] Ling Shao, Fan Zhu, and Xuelong Li. Transfer learning for visual categorization: A survey. *IEEE transactions on neural networks and learning systems*, 26(5):1019–1034, 2014. 4

[131] Astuti Sharma, Tarun Kalluri, and Manmohan Chandraker. Instance level affinity-based transfer for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5361–5371, 2021. 2, 4, 6

[132] Hidetoshi Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of statistical planning and inference*, 90(2):227–244, 2000. 2, 3

[133] Zheng Shou, Dongang Wang, and Shih-Fu Chang. Temporal action localization in untrimmed videos via multi-stage cnns. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1049–1058, 2016. 1

[134] Rui Shu, Hung H. Bui, Hirokazu Narui, and Stefano Ermon. A DIRT-T approach to unsupervised domain adaptation. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net, 2018. 4

[135] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. 1

[136] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014. 1

[137] Baochen Sun, Jiashi Feng, and Kate Saenko. Return of frustratingly easy domain adaptation. *arXiv preprint arXiv:1511.05547*, 2015. 5

[138] Baochen Sun, Jiashi Feng, and Kate Saenko. Correlation alignment for unsupervised domain adaptation. In *Domain Adaptation in Computer Vision Applications*, pages 153–171. Springer, 2017. 2, 4, 5

[139] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision*, pages 443–450. Springer, 2016. 2, 4, 5

[140] Yu Sun, Eric Tzeng, Trevor Darrell, and Alexei A Efros. Unsupervised domain adaptation through self-supervision. *arXiv preprint arXiv:1909.11825*, 2019. 7

[141] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Computer Vision and Pattern Recognition (CVPR)*, 2015. 1

[142] Ilya O Tolstikhin, Bharath K Sriperumbudur, and Bernhard Schölkopf. Minimax estimation of maximum mean discrepancy with radial kernels. *Advances in Neural Information Processing Systems*, 29:1930–1938, 2016. 5

[143] Antonio Torralba and Alexei A Efros. Unbiased look at dataset bias. In *CVPR 2011*, pages 1521–1528. IEEE, 2011. 2

[144] Du Tran, Lubomir Bourdev, Rob Fergus, Lorenzo Torresani, and Manohar Paluri. Learning spatiotemporal features with 3d convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 4489–4497, 2015. 1

[145] Y.-H. Tsai, W.-C. Hung, S. Schulter, K. Sohn, M.-H. Yang, and M. Chandraker. Learning to adapt structured output space for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 4

[146] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7167–7176, 2017. 2, 4, 6

[147] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474*, 2014. 2, 4, 5

[148] Aäron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *CoRR*, abs/1807.03748, 2018. 4

[149] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017. 1

[150] Paul Voigtlaender and Bastian Leibe. Online adaptation of convolutional neural networks for video object segmentation. *arXiv preprint arXiv:1706.09364*, 2017. 1

[151] Tuan-Hung Vu, Himalaya Jain, Maxime Bucher, Matthieu Cord, and Patrick Pérez. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2517–2526, 2019. 4, 7

[152] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. *arXiv preprint arXiv:2006.10726*, 2020. 4, 11

[153] Jindong Wang, Wenjie Feng, Yiqiang Chen, Han Yu, Meiyu Huang, and Philip S Yu. Visual domain adaptation with manifold embedded distribution alignment. In *Proceedings of the 26th ACM international conference on Multimedia*, pages 402–410, 2018. 4, 5

[154] Qian Wang, Penghui Bu, and Toby P Breckon. Unifying unsupervised domain adaptation and zero-shot visual recognition. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2019. 4

[155] Sinan Wang, Xinyang Chen, Yunbo Wang, Mingsheng Long, and Jianmin Wang. Progressive adversarial networks for fine-grained domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9213–9222, 2020. 4, 6

[156] Jason Weston, Antoine Bordes, Sumit Chopra, Alexander M Rush, Bart van Merriënboer, Armand Joulin, and Tomas Mikolov. Towards ai-complete question answering: A set of prerequisite toy tasks. *arXiv preprint arXiv:1502.05698*, 2015. 1

[157] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, abs/1609.08144, 2016. 1

[158] Liang Xiao, Jiaolong Xu, Dawei Zhao, Zhiyu Wang, Li Wang, Yiming Nie, and Bin Dai. Self-supervised domain adaptation with consistency training. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 6874–6880. IEEE, 2021. 2, 4, 7

[159] Saining Xie, Chen Sun, Jonathan Huang, Zhuowen Tu, and Kevin Murphy. Rethinking spatiotemporal feature learning: Speed-accuracy trade-offs in video classification. In *Proceedings of the European conference on computer vision (ECCV)*, pages 305–321, 2018. 1

[160] Shaoan Xie, Zibin Zheng, Liang Chen, and Chuan Chen. Learning semantic representations for unsupervised domain adaptation. In *International Conference on Machine Learning*, pages 5423–5432, 2018. 4, 6

[161] Jiaolong Xu, Liang Xiao, and Antonio M López. Self-supervised domain adaptation for computer vision tasks. *IEEE Access*, 7:156694–156706, 2019. 2, 4, 7

[162] Ruijia Xu, Guanbin Li, Jihan Yang, and Liang Lin. Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation. In *Proceedings*

of the IEEE International Conference on Computer Vision, pages 1426–1435, 2019. 4, 6

[163] Hongliang Yan, Yukang Ding, Peihua Li, Qilong Wang, Yong Xu, and Wangmeng Zuo. Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2272–2281, 2017. 4, 5

[164] Seunghan Yang, Youngeun Kim, Dongki Jung, and Changick Kim. Partial domain adaptation using graph convolutional networks. *ArXiv*, abs/2005.07858, 2020. 4, 8

[165] Shiqi Yang, Yaxing Wang, Joost van de Weijer, and Luis Herranz. Unsupervised domain adaptation without source data by casting a bait. *ArXiv*, abs/2010.12427, 2020. 4, 11

[166] Shiqi Yang, Yaxing Wang, Joost van de Weijer, Luis Herranz, and Shangling Jui. Generalized source-free domain adaptation. *arXiv preprint arXiv:2108.01614*, 2021. 4, 11

[167] Kaichao You, Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Universal domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2720–2729, 2019. 4, 9, 10, 11

[168] Xiangyu Yue, Zangwei Zheng, Shanghang Zhang, Yang Gao, Trevor Darrell, Kurt Keutzer, and Alberto Sangiovanni Vincentelli. Prototypical cross-domain self-supervised learning for few-shot unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13834–13844, 2021. 4

[169] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny. Barlow twins: Self-supervised learning via redundancy reduction. In Marina Meila and Tong Zhang, editors, *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*, pages 12310–12320. PMLR, 2021. 4

[170] Jing Zhang, Zewei Ding, Wanqing Li, and Philip Ogunbona. Importance weighted adversarial nets for partial domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8156–8164, 2018. 4, 8, 10

[171] Jing Zhang, Wanqing Li, and Philip Ogunbona. Joint geometrical and statistical alignment for visual domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1859–1867, 2017. 2, 4, 5

[172] Richard Zhang, Phillip Isola, and Alexei A Efros. Colorful image colorization. In *European conference on computer vision*, pages 649–666. Springer, 2016. 4

[173] Weichen Zhang, Wanli Ouyang, Wen Li, and Dong Xu. Collaborative and adversarial network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3801–3809, 2018. 2, 4

[174] An Zhao, Mingyu Ding, Zhiwu Lu, Tao Xiang, Yulei Niu, Jiechao Guan, and Ji-Rong Wen. Domain-adaptive few-shot learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1390–1399, 2021. 4, 7

[175] Erheng Zhong, Wei Fan, Jing Peng, Kun Zhang, Jiangtao Ren, Deepak Turaga, and Olivier Verscheure. Cross domain distribution adaptation via kernel mapping. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1027–1036, 2009. 2

[176] Jinhua Zhu, Yingce Xia, Lijun Wu, Di He, Tao Qin, Wengang Zhou, Houqiang Li, and Tie-Yan Liu. Incorporating bert into neural machine translation. *arXiv preprint arXiv:2002.06823*, 2020. 1

[177] Weiming Zhuang, Xin Gan, Yonggang Wen, Xuesen Zhang, Shuai Zhang, and Shuai Yi. Towards unsupervised domain adaptation for deep face recognition under privacy constraints via federated learning. *arXiv preprint arXiv:2105.07606*, 2021. 2

[178] Fariba Zohrizadeh, Mohsen Kheirandishfard, and F. Kamangar. Class subset selection for partial domain adaptation. In *CVPR Workshops*, 2019. 4, 8