Adaptive Adversarial Network for Source-free Domain Adaptation

Haifeng Xia[†], Handong Zhao[‡], and Zhengming Ding[†]

[†]Department of Computer Science, Tulane University, USA

[‡]Adobe Research, San Jose, USA

hxia@tulane.edu, hazhao@adobe.com, zding1@tulane.edu

Abstract

*Unsupervised Domain Adaptation solves knowledge transfer along with the coexistence of well-annotated source domain and unlabeled target instances. However, the source domain in many practical applications is not always accessible due to data privacy or the insufficient memory storage for small devices. This scenario defined as **Source-free Domain Adaptation** only allows accessing the well-trained source model for target learning. To address the challenge of source data unavailability, we develop an Adaptive Adversarial Network (**A²Net**) including three components. Specifically, the first one named Adaptive Adversarial Inference seeks a target-specific classifier to advance the recognition of samples which the provided source-specific classifier difficultly identifies. Then, the Contrastive Category-wise Matching module exploits the positive relation of every two target images to enforce the compactness of subspace for each category. Thirdly, Self-Supervised Rotation facilitates the model to learn additional semantics from target images by themselves. Extensive experiments on the popular cross-domain benchmarks verify the effectiveness of our proposed model on solving adaptation task without any source data.*

1. Introduction

Recent years witness great promising achievements from the exploration of deep neural network (DNN) in the practical scenarios, i.e., image classification, segmentation and detection [20, 47, 24, 48]. However, DNN model easily suffers from severe performance degradation when evaluated on test data (target domain) lying in different distribution from the training instances (source domain). Such discrepancy termed as domain shift [9, 36] results from the varying environments or devices [28] and various image styles [49]. To tackle the challenge, unsupervised domain adaptation (UDA) attracts increasing attention and achieves encouraging results by using deep learning architecture.
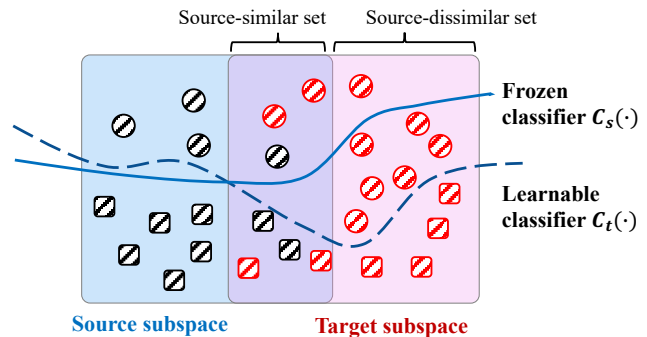
Conventional UDA assumes the cross-domain data is

Figure 1. Schematic diagram of high-level source (black) and target (red) feature distributions from the trained source model. The target samples can be divided into two subsets: source-similar and source-dissimilar sets. Square and circle represent two different categories. **A²Net** adaptively learns a new classifier (dashed) based on the frozen classifier (solid) trained in source domain.

available during model training, so that it effectively measures and eliminates the domain discrepancy [7, 39, 42]. Based on this assumption, the mainstream solutions of UDA are roughly divided into two paradigms. One branch attempts to transform source and target data into the high-level feature space with the consistent statistical moments to achieve the alignment of their feature distributions [14, 26, 41, 3, 16]. As for the representative work maximum mean discrepancy (MMD), the learned cross-domain features are enforced to share the identical first-order moment. The other branch devotes more efforts to the deployment of adversarial framework [27, 33, 2]. The core strategy is exploiting feature generator to deceive the domain discriminator so that it fails to recognize which domain the feature comes from. Despite the successes of these methods, it is not hard to observe that they heavily depend on the co-existence of source and target data. However, abundant application scenarios cannot always meet the basic assumption of UDA due to data privacy and memory constraint of small devices. For instance, the training benchmark of ImageNet [6] contains 14 million images occupying hundreds gigabytes storage, which is a huge burden for small-storage equipment. Moreover, many industries such as hospitals are

restricted to share their sensitive data with external sites.

The conflict between the practical demand and UDA setting motivates the novel research direction named *Source-free Domain Adaptation* where we are only provided with the well-trained source model instead of well-annotated source data to achieve adaptation to target data. Recently, a few research efforts [21, 19] start exploring this new scenario on cross-domain classification task by assuming that the source classifier contains sufficient knowledge. Thus, they both attempt to directly adjust target features to adapt the source classifier. Among them, SHOT [21], as a simple yet efficient method, freezes the source classifier and integrates pseudo-label supervision and entropy minimization [46] to shorten the distance between target features and source classification boundary. Similarly, MA [19] first considers source classifier as an anchor to guide the generation of new target samples closer to the source domain and then adopts adversarial strategy to achieve domain alignment. In addition, SoFA [43] adopts self-supervised reconstruction to extract more discriminative knowledge from target images themselves to improve the classification ability of model. However, when the data in source domain is imbalanced or insufficient, the above methods with frozen classifier becomes vulnerable due to the lower generalization of source classifier. It is difficult for these approaches to move abundant target features with large variance into the small source classification boundary. For example, as illustrated in Figure 1, the source classifier (solid line) trained on the imbalanced data where circle class has only a few data points. Restricted by the frozen classifier, this, unfortunately, leads to a bad classification performance in source-dissimilar set. From another perspective, we post a question: "Can we seek a novel target-specific classifier during model optimization and adapt it to the target features?".

Along with such a question, we propose a novel Adaptive Adversarial Network ($\mathbf{A^2Net}$) to address the Source-Free Domain Adaptation. To achieve flexible adjustment for classifier and preserve the original source knowledge, our work firstly introduces a novel target classifier and then exploits dual-classifier design to achieve adversarial domain-level alignment and contrastive category-wise matching (CCM). Concretely, according to the predictions of source and target classifiers, we adaptively divide target samples into two categories: source-similar set and source-dissimilar one. By building such an adversarial relation between dual-classifier and feature generator, $\mathbf{A^2Net}$ gradually eliminates the significant difference across source-similar and source-dissimilar sets and remedies the defect of the frozen source classifier by updating the target classifier. To further learn discriminative features, our work considers the relation of paired samples consisting two any target images as three levels: positive, uncertain and negative pairs, and develops contrastive category-wise matching over all

positive pairs to intensify their association. The main contributions of our work are summarized as three folds:

- First, the proposed $\mathbf{A^2Net}$ integrates a new flexible classifier to be available for optimization and the frozen source classifier to form the dual-classifier architecture which we use to adaptively distinguish source-similar target samples from source-dissimilar ones and achieve alignment across them.

- Second, $\mathbf{A^2Net}$ learns robust and discriminative features in a self-supervised learning manner. Specifically, the contrastive category-wise matching module relies on source knowledge to explore the association of the paired target features and enforce the positive relation to achieve category-wise alignment.

- Finally, we further enhance the model to learn additional semantics through a self-supervised rotation. Experimental results on three benchmarks fully verify the effectiveness of $\mathbf{A^2Net}$ for source-free scenario.

2. Related Works

Unsupervised Domain Adaptation (UDA) aims to transfer knowledge of source domain to achieve the classification of unsupervised target domain [12]. And these two domains share the identical label space yet with different distributions due to the varying conditions [13, 35]. During the training stage, the models have the access to well-annotated source data and unlabeled target data. Thus, the intuition is to learn domain-invariant representation with all available data so that the well-trained source model is directly applied to the target data. Recently, many efforts [46, 38] make significant progress on this exploration by designing various distribution alignment criteria. From domain-level alignment, several works [26, 31] attempt to match various order moments across source and target distributions, while others [33, 46] introduce discriminator to achieve adversarial game between feature generator and discriminator. As for the category-wise alignment, CAN [14] exploits the pseudo-label of target sample to reduce the intra-class distance and learn more compact category subspace. In terms of the sample-wise alignment, a few works [18, 5] extract and optimize sample-to-sample association via optimal transport. Although these methods actually have facilitated model with powerful adaptation ability, they heavily rely on availability of two domains' data, which conflicts with the real applications due to the data privacy and memory constraint of small devices. To surmount the bottleneck of UDA, we consider source-free domain adaptation and develop an adaptive adversarial network.

Source-free Domain Adaptation provides the well-trained source model and target data without any access to source data during the training stage. Under this condition, the
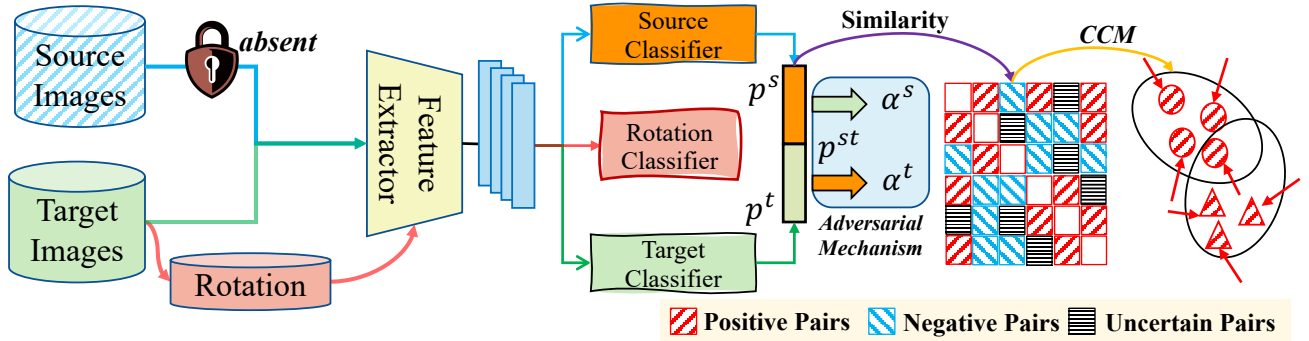
Figure 2. Overview of our Adaptive Adversarial Network ($\mathbf{A^2Net}$) on solving source-free domain adaptation. Given the trained source model including feature extractor $F(\cdot)$ and source classifier $C_s(\cdot)$, we transfer it to identify the target images without source data. To address such a challenge, $\mathbf{A^2Net}$ first adaptively distinguishes source-similar target samples from source-dissimilar ones, and adopts soft-adversarial manner with the introduced target classifier to eliminate their discrepancy. Second, our method explores the contrastive category-wise matching (CCM) to reinforce the relation of positive paired samples. Third, $\mathbf{A^2Net}$ exploits self-supervised rotation to learn more robust and discriminative features.

conventional UDA methods become invalid since they fail to match source feature with target ones. Most recently, [19, 17, 21] discover that the original trained model conceals lots of knowledge of source feature distribution. Thus, with the supervision of source classifier, [19, 17] attempt to produce novel target samples closer to source domain and then align the novel and original target instances in high-level features via adversarial manner. Similarly, [21] freezes the source classifier and applies pseudo-label to optimize feature generator, which aims to move target features into the unseen source feature domain. However, the significant domain discrepancy or imbalanced distribution of source domain has a negative influence on the generalization of source model, which increases the difficulty of adapting target feature to source classifier. Differently, our work introduces a novel flexible target classifier under dual-classifier architecture with the frozen source classifier. The dual-classifier firstly assists to adaptively discover source-similar and source-dissimilar target sets, and fights with the feature generator to achieve the alignment across the two subsets. In addition, we adopt self-supervised manners, i.e., category-wise matching and rotation recognition, to learn more discriminative representations.

3. The Proposed Method

3.1. Preliminaries

Given the well-annotated source domain $\mathscr{D}_s = \{(x_i^s, y_i^s)\}$ and unlabeled target instances $\mathscr{D}_t = \{x_i^t\}$, the conventional UDA methods [40, 23, 45, 34] attempt to eliminate domain discrepancy by training a model with the available cross-domain data. However, the practical applications sometimes are restricted to access to original source raw data due to data privacy and/or memory constraint of small devices, which motivates the more challenging Source-Free Domain Adaptation. When adapting to target domain in the

novel scenario, we only deploy well-trained source model including feature extractor $F(\cdot)$ and classifier $C_s(\cdot)$ to recognize target samples without any source instances for explicit cross-domain alignment. With the main exploration on how to adapt model to target classification task, our work follows the protocol [21] to train source model by optimizing the $F(\cdot)$ and $C_s(\cdot)$ with the supervisor of source annotation and neglect the specific description on this part.

3.2. $\mathbf{A^2Net}$: Adaptive Adversarial Network

From the investigation of Figure 1, the considerable domain discrepancy results in the mismatch of feature distribution across source and target domains. Fortunately, there exist some ready-to-recognize target instances similar to source domain distribution for each category. Thus, target high-level features can be divided into two types: source-similar features and source-dissimilar ones. The source classifier $C_s(\cdot)$ confidently identifies source-similar samples. However, it difficultly provides accurate labels for the remaining, especially when trained on insufficient data in source domain. Under such condition, the frozen source classifier in SHOT [21] becomes invalid for the classification of source-dissimilar features, since it is difficult to adapt abundant target features with large variance to $C_s(\cdot)$ with the lower generalization. To avoid the defect, we propose a novel method named Adaptive Adversarial Network ($\mathbf{A^2Net}$) in Fig. 2 which alternatively develops a learnable classifier $C_t(\cdot)$ to adapt target feature distribution. The introduced target classifier not only should accurately identify source-similar target feature as C_s but also improves the recognition ability on source-dissimilar ones. Along with the mentioned expectation, the first challenge is to distinguish source-similar features from source-dissimilar ones. However, it is non-trivial to make the decision due to the difficulty of measuring distance between data points and class boundary in high-dimensional feature space.

3.2.1 Soft-Adversarial Inference

Motivated by the voting strategy, we compare the output of classifiers to adaptively determine the type of features. Specifically, each target sample through two classifiers in Figure 2 achieves its probability distribution of category before Softmax operation $p_i^s = C_s(F(x_i^t)) \in \mathbb{R}^K$ and $p_i^t = C_t(F(x_i^t)) \in \mathbb{R}^K$, where K is the number of class. Subsequently, we activate the concatenation of p_i^s and p_i^t with Softmax function $\sigma(\cdot)$ to access $p_{(i)}^{st} = \sigma([p_i^s\ p_i^t]^\top) \in \mathbb{R}^{2K}$, and consider $\alpha_i^s = \sum_{k=1}^K p_{(i)k}^{st}$ and $\alpha_i^t = \sum_{k=K+1}^{2K} p_{(i)k}^{st}$ as voting scores. When α_i^s is larger than α_i^t, the corresponding feature belongs to source-similar set, otherwise, it is divided into the other subset. The definition gives us a manner to optimize target classifier and feature extractor with the following formulation:

$$
\begin{aligned}
\min_{F,C_t} \quad & -\sum_{i=1}^{n_t} \mathbb{I}(\alpha_i^s > \alpha_i^t)\sigma(p_i^s)\log(\sigma(p_i^s)) \\
& -\sum_{i=1}^{n_t} \mathbb{I}(\alpha_i^s \le \alpha_i^t)\sigma(p_i^t)\log(\sigma(p_i^t)),
\end{aligned} \quad (1)
$$

where $\mathbb{I}(\cdot)$ is the indicator function. However, such a constraint easily gives rise to the necessary concern "What will happen if $C_t(\cdot)$ generates wrong prediction when $\alpha_i^s \le \alpha_i^t$?" Under this situation, the prediction tends to be far away from the ground-truth. Thus, the trade-off between accepting novel target knowledge and preserving well-learned source knowledge becomes important, and we further rewrite Eq. (1) as:

$$
\mathcal{L}_c = -\sum_{i=1}^{n_t} \Big(\alpha_i^s \sigma(p_i^s)\log\big(\sigma(p_i^s)\big) + \alpha_i^t \sigma(p_i^t)\log\big(\sigma(p_i^t)\big) \Big),
$$

where α_i^s and α_i^t are frozen during optimization.

From another perspective, we also consider the source-similar and source-dissimilar high-level features distributing in two independent domains. The alignment of them further reduces their discrepancy to learn more discriminative features. In addition, the introduced target classifier $C_t(\cdot)$ finally has the equivalent classification ability for source-similar ones. According to the dual-classifier design, we propose a **Soft-Adversarial** mechanism to address the above demands with the formal objective function as:

$$
\min_{C_t} \mathcal{L}_{c'} = -\sum_{i=1}^{n_t}(\alpha_i^s \log(\sum_{k=1}^K p_{(i)k}^{st}) + \alpha_i^t \log(\sum_{k=K+1}^{2K} p_{(i)k}^{st})),
$$

$$
\min_{F} \mathcal{L}_{c''} = -\sum_{i=1}^{n_t}(\alpha_i^t \log(\sum_{k=1}^K p_{(i)k}^{st}) + \alpha_i^s \log(\sum_{k=K+1}^{2K} p_{(i)k}^{st})).
$$

To explicitly understand the Soft-Adversarial loss, we firstly illustrate that $\alpha_i^{s/t}$ denotes the probability of the sample x_i belonging to source-similar or source-dissimilar subsets and $\alpha_i^s + \alpha_i^t = 1$. For the extreme condition such as

$\alpha_i^s \approx 1$, the optimization of $\mathcal{L}_{c'}$ further reduces the discriminability of target classifier $C_t(\cdot)$ for x_i. However, feature generator engages in the inverse operation mapping x_i into high-level representation similar to source-dissimilar part by minimizing $\mathcal{L}_{c''}$. Beneficial from the adversarial manner between feature generator and classifiers, we further align source-closer and source-dissimilar sets and eliminate the difference of classifiers.

3.2.2 Contrastive Category-wise Matching

The core motivation of adaptive adversarial inference is to discover source-similar target samples and achieve domain-level alignment across source-similar and source-dissimilar sets. However, domain-invariant features learned with adversarial learning fail to represent the category-level matching. In addition, without annotation over target domain, it becomes difficult to identify the category relationship among samples. The intuitive solution to the challenge is to directly provide each sample with pseudo-label, which easily results in the negative influence on model training, especially during the initialization stage, due to the uncertainty of pseudo-label. Inspired by the contrastive learning [1], we design a novel discriminative dual classifier exploring the association of paired samples to achieve the class-wise alignment in unsupervised manner.

Concretely, each visual instance within a batch is transformed into the label space via the source classifier $\mathcal{G}_i = \sigma(p_i^s) \in \mathbb{R}^K$ which we use to capture the similarity of any paired samples through $s_{ij} = \mathcal{G}_i^\top \mathcal{G}_j$ in Figure 2. The larger s_{ij} denotes that the i-th and j-th samples belong to the same category with higher probability, vice versa. However, we fail to confidently judge the relationship of several pairs when s_{ij} lies in the middle interval. Thus, all pairs of each mini-batch are divided into three subsets: positive, uncertain and negative sets by comparing s_{ij} with the upper bound $\mu(t)$ and lower bound $\ell(t)$ defined as:

$$
\begin{cases}
\mu(t) = \mu_0 - \lambda_\mu \cdot t \\
\ell(t) = \ell_0 + \lambda_\ell \cdot t \\
0 \le \ell(t) \le \mu(t) \le 1
\end{cases}
\quad
\gamma_{ij} =
\begin{cases}
1, & s_{ij} > \mu(t) \\
-1, & s_{ij} < \ell(t) \\
0, & otherwise
\end{cases}
$$

where $\mu(t)$ as well as $\ell(t)$ are the linear functions of epoch t starting from zero, and μ_0 and ℓ_0 are the initial upper and lower bounds, respectively, and λ_μ and λ_ℓ separately control the decreasing and increasing rate of μ_0 and ℓ_0. The pairs are definitely classified into positive ($\gamma_{ij} = 1$) and negative ($\gamma_{ij} = -1$) subsets when $s_{ij} > \mu(t)$ and $s_{ij} < \ell(t)$, respectively. For other cases, we temporarily neglect the ambiguous associations with $\ell(t) < s_{ij} < \mu(t)$ by $\gamma_{ij} = 0$. As the piecemeal change of $\mu(t)$ and $\ell(t)$, our method adaptively makes the judgement for more pairs.

To achieve class-wise alignment, the relation of positive pairs must be further intensified to learn more similar fea-

ture representation for themselves. Similar with [1], we develop the contrastive loss for each positive pair of example (i, j) formulated as:

$$\xi(i,j) = -\log \frac{\exp(s_{ij})}{\sum_{v=1}^{b} \mathbb{I}(v \neq i)|\gamma_{iv}|\exp(s_{iv})}, \quad (2)$$

where b is the size of each batch and $|\gamma_{iv}|$ means the absolute value of γ_{iv}. According to the monotonic property[1], we obtain the optimization of Eq. (2) approximates the minimum value of function with $s_{ij} \to 1$. That illustrates $\sigma(p_i^s)$ and $\sigma(p_j^s)$ follow the more similar probability distribution. And the property is transmitted into the output of feature generator due to the frozen source classifier so that samples from the identical class distribute closer to each other in the high-level feature space. Thus, we adopt Eq. 2 on all positive pairs and reformulate the final contrastive objective:

$$\min_{F} \mathcal{L}_p = \mathbb{I}[\mu(\lambda) > \ell(\lambda)] \sum_{i=1}^{b} \sum_{j=1, j \neq i}^{b} \mathbb{I}(\gamma_{ij} = 1)\xi(i,j). \quad (3)$$

Note that Eq. (3) makes no sense under $\mu(\lambda) \leq \ell(\lambda)$ since we fail to find new positive pairs to optimize it.

3.2.3 Self-Supervised Rotation

So far, we mainly consider how to transfer knowledge into target domain only with the guidance of well-trained source model. However, the pure classification model heavily relies on the given source data, which is often lying imbalanced distribution, further limiting the generalization ability of target classifier. To solve this, we explore self-supervised rotation manner over target domain to augment the sample space, which enhances the learning of feature extraction and target classifier. In other words, the model is able to easily see more variances per high-confident predicted target sample. Following [44], we set four rotation degrees $\theta \in \{0^o, 90^o, 180^o, 270^o\}$ with corresponding 4-class rotation labels y_r. Within one batch, we randomly select rotation label y_r and then have access to the new processed image $\hat{x}_i^t$ by rotating the original image x_i^t with $90^o y_r$. In addition, we also introduce the rotation classifier $C_r(\cdot)$ in Figure 2 taking $F(\hat{x}_i^t)$ as input and predicting the rotation label. Finally, the cross-entropy loss is adopted to measure the difference between prediction and ground-truth rotation as follows:

$$\min_{F, C_r} \mathcal{L}_r = -\sum_{i=1}^{b} y_i^r \log(F(\hat{x}_i^t)). \quad (4)$$

---

[1]When the variable $x \in [0,1]$, the objective function $f(x) = -\log \frac{\exp(x)}{\exp(x)+a}$ achieves the minimum value when $x = 1$, since $f'(x) < 0$ meaning $f(x)$ is monotonically decreasing when $x \in [0,1]$.

By identifying the rotation degree, the model effectively captures the important visual signals from original images for object classification.

3.3. Overall Objective and Optimization

The above description has specifically illustrated how our method works for source-free domain adaptation. It is simple to notice that the training of model mainly involves the update of three modules (i.e., feature generator $F(\cdot)$, rotation classifier $C_r(\cdot)$, and target classifier $C_t(\cdot)$) with the following overall objective:

$$\min_{F, C_r} \mathcal{L}_c + \mathcal{L}_{c''} + \mathcal{L}_p + \eta \mathcal{L}_r, \quad (5)$$

$$\min_{C_t} \mathcal{L}_c + \mathcal{L}_{c'}, \quad (6)$$

where η is the trade-off parameter. To achieve the adaptive adversarial operation, we adopt iterative manner to alternately optimize three modules. **First**, the source and target classifiers take the features from generator as input to access the class prediction which we use to update the feature generator and rotation classifier via Eq. (5) with frozen the target multi-class classifier $C_t(\cdot)$. **Second**, we only optimize the target classifier when fixing $F(\cdot)$ and $C_r(\cdot)$ with Eq. (6). **Third**, the adversarial training repeats the above two steps until we reach the convergence or maximum epochs.

4. Experiments

4.1. Experimental Details

§ **Datasets.** *Office-31* [30] as a popular benchmark contains three domains: Amazon (**A**), Webcam (**W**) and DSLR (**D**). The specific number of image for each domain is 2,817 (**A**), 795 (**W**) and 498 (**D**), and three domains share the identical label space with 31 categories. Adapting model from source to target domain on Office-31 not only eliminates the domain shift but also overcomes additional challenge resulting from the data scale imbalance such as task **A**→**D**. *Office-Home* [34] consists of 15,500 images collected from four domains Realworld (**Rw**), Clipart (**Cl**), Art (**Ar**), Product (**Pr**) and there are 65 categories per domain. The challenges of 12 domain adaptation tasks mainly result from more object categories and considerable domain discrepancy. *VisDA* [29] is a challenging large-scale benchmark with 12 classes typically used to evaluate model adaptation from synthetic domain to real domain. The source domain involves 152 thousands synthetic images which the 3D rendering model produces under various conditions. The validation set is considered as the target domain containing 55 thousands real object images from MS-COCO [22].

§ **Implementation.** According to [21], for the source model, we separately consider Resnet-50 and Resnet-101 as backbones to extract high-level features from two object

Table 1. Comparisons of Object Classification Accuracy (%) of Source-free Domain Adaptation on Office-31. The best accuracy for source-free tasks is highlighted with **bold** type, while we use <u>underline</u> to emphasize the highest result for source-need task.

	Method	A→D	A→W	D→A	D→W	W→A	W→D	Avg
Source-Needed	ResNet [10]	68.9	68.4	62.5	96.7	60.7	99.3	76.1
	DANN [8]	79.7	82.0	68.2	96.9	67.4	99.1	82.2
	SAFN [38]	90.7	90.1	73.0	98.6	70.2	99.8	87.1
	CDAN [25]	92.9	94.1	71.0	98.6	69.3	100.0	87.7
	BNM [4]	90.3	91.5	70.9	98.5	71.6	100.0	87.1
	MCC [11]	95.6	95.4	72.6	98.6	73.9	100.0	89.4
	SRDC [32]	<u>95.8</u>	<u>95.7</u>	<u>76.7</u>	<u>99.2</u>	<u>77.1</u>	100.0	<u>90.8</u>
Source-Free	SFDA [15]	92.2	91.1	71.0	98.2	71.2	99.5	87.2
	SDDA [17]	85.3	82.5	66.4	99.0	67.7	99.8	83.5
	SoFA [43]	73.9	71.7	53.7	96.7	54.6	98.2	74.8
	SHOT [21]	94.0	90.1	74.7	98.4	74.3	99.9	88.6
	Ours	**94.5**	**94.0**	**76.7**	**99.2**	**76.1**	**100.0**	**90.1**

datasets and VisDA, and replace the original last FC layer with a new bottleneck layer followed by Batch Normalization (BN). The source classifier C_s consists of one FC layer and a weight normalization layer. During adaptation, we introduce the target classifier C_t with the same architecture as C_s and the rotation classifier C_r including two FC layers. In addition, we initialize C_t with the parameters of C_s by appending Gaussian noises from $N(0, I)$. As for the optimizer, we adopt SDG with momentum 0.9 and weight decay $1e^{-3}$. The initial learning rates on Office-31/Office-Home for the pre-trained backbone and new added components are $1e^{-3}$ and $1e^{-2}$ respectively, however, they are set as $1e^{-4}$ and $1e^{-2}$ for VisDA. Moreover, we set the identical upper and lower bounds for all experiments as $\mu_0 = 0.95$, $\ell_0 = 0.45$, $\lambda_\mu = 9.9e^{-3}$ and $\lambda_\ell = 9.9e^{-4}$. Source code of this work is available online[2].

§ **Baselines.** The comparisons include two categories of domain adaptation algorithms. One is vanilla domain adaptation, which requires source and target data at the same time to solve the domain shift, such as Resnet [10], DANN [8], SAFN [38], CDAN [25], SRDC [32], BNM [4] and MCC [11]. Additionally, we also compare the recent state-of-the-art source-free domain adaptation models, i.e., SFDA [15], SHOT [21], SDDA [17] and SoFA [43]. Note that since MA [19] needs to generate additional target samples on solving source-free task, we make no comparison with it.

4.2. Comparison Results

Tables 1-3 report the results of object classification on Office-31, Office-Home and VisDA, respectively. From the investigation of them, our proposed Adaptive Adversarial Network ($\mathbf{A^2Net}$) achieves the highest average accuracy across three benchmarks when compared with others

---

for source-free domain adaptation, which illustrates the design of $\mathbf{A^2Net}$ effectively transfers knowledge only from source model to assist the target data recognition. In addition, we can easily achieve three important conclusions by making delicate comparisons over these competitors.

First of all, $\mathbf{A^2Net}$ provides well-trained source model with more powerful adaptation ability when evaluated on unsupervised target domain, especially for small scale source domain. For example, with **D** and **W** as source domains on Office-31, our approach outperforms the second highest accuracy from SHOT by 2.0% and 1.8% when adapting to the target domain **A**. As we all know, there exists serious imbalanced data scale challenge across source and target domain, i.e., **D** (498) vs **A** (2,817) and **W** (795) vs **A**. The classifier firstly trained on small-scale source domain has so insufficient generalization ability that it ineffectively is applied to large-scale target domain. Thus, it is difficult for SHOT with frozen classifier to accurately move abundant target features into the source classification boundary. However, our $\mathbf{A^2Net}$ adopts flexible target classifier with adversarial training to adapt it to target features. This is the main reason for our success on these two tasks. And $\mathbf{A^2Net}$ beats several UDA based methods such as SAFN and BNM by a large margin, which means even if we fail to access the source data, our method still exploits the finite source knowledge to achieve better adaptation.

Second, our proposed method also effectively overcomes the negative influence of significant domain discrepancy. To the best of our knowledge, there exists significant domain shift between **Ar** and **Cl** because of the considerable difference of image styles. However, $\mathbf{A^2Net}$ surpasses SFDA by 10% for this adaptation task since our proposed method adaptively distinguishes the source-similar target samples from source-dissimilar ones and explores adversarial manner to gradually eliminate domain discrepancy. Moreover, with the increasing of object category, we notice that all methods suffer from the performance degradation on Office-Home when compared with their results of Office-31. But the contrastive category-wise matching depends on the constraint over positive paired samples to learn so explicit classification boundary that $\mathbf{A^2Net}$ still achieves the best classification accuracy among the baselines.

Third, the experimental results in Table 3 fully demonstrate that our designed algorithm makes sense to solve source-free domain adaptation with the large-scale benchmark. Specifically, $\mathbf{A^2Net}$ obtains higher classification accuracy than other state-of-the-art methods in most adaptation tasks on VisDA and makes more accurate identify on several confusing objects such as bus and car.

4.3. Empirical Analysis

§ **Feature Visualization & Confusion Matrix.** According to the experimental results in Table 1 and the working

Table 2. Comparisons of Object Classification Accuracy (%) of Source-free Domain Adaptation on Office-Home. The best accuracy for source-free tasks is highlighted with **bold** type, while we use underline to emphasize the highest result for source-need task.

	Method	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	Avg
Source-Needed	Resnet[10]	46.3	67.5	75.9	59.1	59.9	62.7	58.2	41.8	74.9	67.4	48.2	74.2	61.3
	DANN[8]	45.6	59.3	70.1	47.0	58.5	60.9	46.1	43.7	68.5	63.2	51.8	76.8	57.6
	SAFN[37]	52.0	71.7	76.3	64.2	69.9	71.9	63.7	51.4	77.1	70.9	57.1	81.5	67.3
	CDAN[25]	50.7	70.6	76.0	57.6	70.0	70.0	57.4	50.9	77.3	70.9	56.7	81.6	65.8
	BNM[4]	52.3	73.9	80.0	63.3	72.9	74.9	61.7	49.5	79.7	70.5	53.6	82.2	67.9
	SRDC[32]	52.3	76.3	81.0	69.5	76.2	78.0	68.7	53.8	81.7	76.3	57.1	85.0	71.3
Source-Free	SFDA[15]	48.4	73.4	76.9	64.3	69.8	71.7	62.7	45.3	76.6	69.8	50.5	79	65.7
	SoFA[43]	-	74.1	77.6	-	71.8	75.1	-	-	-	-	-	-	-
	SHOT[21]	57.1	78.1	81.5	**68.0**	78.2	78.1	67.4	54.9	82.2	73.3	58.8	84.3	71.8
	Ours	**58.4**	**79.0**	**82.4**	67.5	**79.3**	**78.9**	**68.0**	**56.2**	**82.9**	**74.1**	**60.5**	**85.0**	**72.8**

Table 3. Comparisons of Object Classification Accuracy (%) of Source-free Domain Adaptation on VisDA. The best accuracy for source-free tasks is highlighted with **bold** type, while we use underline to emphasize the highest result for source-need task.

	Methods	plane	bcycl	bus	car	horse	knife	mcycl	person	plant	sktbrd	train	truck	Per-Class
Source-Needed	Resnet [10]	55.1	53.3	61.9	59.1	80.6	17.9	79.7	31.2	81.0	26.5	73.5	8.5	52.4
	DANN [8]	81.9	77.7	82.8	44.3	81.2	29.5	65.1	28.6	51.9	54.6	82.8	7.8	57.4
	CDAN [25]	85.2	66.9	83.0	50.8	84.2	74.9	88.1	74.5	83.4	76.0	81.9	28.0	73.9
	SAFN [38]	93.6	61.3	84.1	70.6	94.1	79.0	91.8	79.6	89.9	55.6	89.0	24.4	76.1
	MCC [11]	88.7	80.3	80.5	71.5	90.1	93.2	85.0	71.6	89.4	73.8	85.0	36.9	78.8
Source-Free	SFDA [15]	86.9	81.7	84.6	63.9	93.1	91.4	**86.6**	71.9	84.5	58.2	74.5	42.7	76.7
	SoFA [43]	-	-	-	-	-	-	-	-	-	-	-	-	64.6
	SHOT [21]	**94.3**	**88.5**	80.1	57.3	93.1	94.9	80.7	80.3	91.5	**89.1**	86.3	**58.2**	82.9
	Ours	94.0	87.8	**85.6**	**66.8**	93.7	**95.1**	85.8	**81.2**	**91.6**	88.2	**86.5**	56.0	**84.3**

mechanism of model, when compared with other state-of-the-art baselines, it is simple to notice that A^2Net is non-sensitive to the mismatch of cross-domain data scale, where source domain contains much fewer instances than target domain. To further explore how our work achieves it, we provide the visualization of embedding feature and confusion matrix in Figure 3 with large- or small- scale source domain. Concretely, the well-trained target model of A^2Net and source-only ResNet are frozen to extract the high-level features before the classifier from the unseen source domain and unlabeled target one. And we carry out the experiments on Office-31 since it exists the imbalanced data scale challenge, i.e., **A** (2,817) vs **D** (498), and **A** (2,817) vs **W** (795). The comparison between Fig. 3 (a) and (b) illustrates the model trained on large-scale source domain has more powerful generalization ability than that with smaller-scale one. With **A** as source domain, A^2Net easily distinguishes source-similar target features from source-dissimilar ones and gradually aligns these two parts by using soft-adversarial mechanism. Thus, after adaptation, target features of each category (produced by A^2Net) almost distribute the boundary of source domain. Under this condition, our target classifier being similar to the original source one exactly identifies them. However, with source model trained on **D**, even if the model has finished the adaptation, there are abundant target features far away from the corre-sponding source class so that the original source classifier difficultly make an accurate decision on them. The flexible target classifier of our method, thus, fully shows the importance of its optimization, which facilitates model to adapt itself to target features instead of only adjusting feature learning as SHOT [21]. Beneficial from the dual-classifier design, our work achieves the highest classification accuracy on task **D→A** in Table 1. In addition, the confusion matrix derived from task **W→A** demonstrates our method learns more compact category subspace by intensifying the association of positive pairs with contrastive loss to achieve category-wise matching across source-similar and source-dissimilar sets.

§ **Ablation Study, Parameter Analysis & Training Stability.** Our A^2Net mainly consists of three modules: soft-adversarial inference, contrastive category-wise matching and self-supervised rotation which support the model adaptation from various perspectives. Therefore, we attempt to separately remove each module from them to investigate the change of classification accuracy on two tasks **D→A** and **W→A**. According to the experimental results in Fig. 4 (b), we achieve the conclusion that the soft-adversarial mechanism has an important and positive influence on improving the generalization of model. Without the adversarial operation, it becomes tough to effectively promote adaptation of the target classifier so that the model heavily relies on the

(a) ResNet $(A \rightarrow W)$ (b) $\mathbf{A^2Net}(A \rightarrow W)$ (c) ResNet $(D \rightarrow A)$ (d) $\mathbf{A^2Net}(D \rightarrow A)$

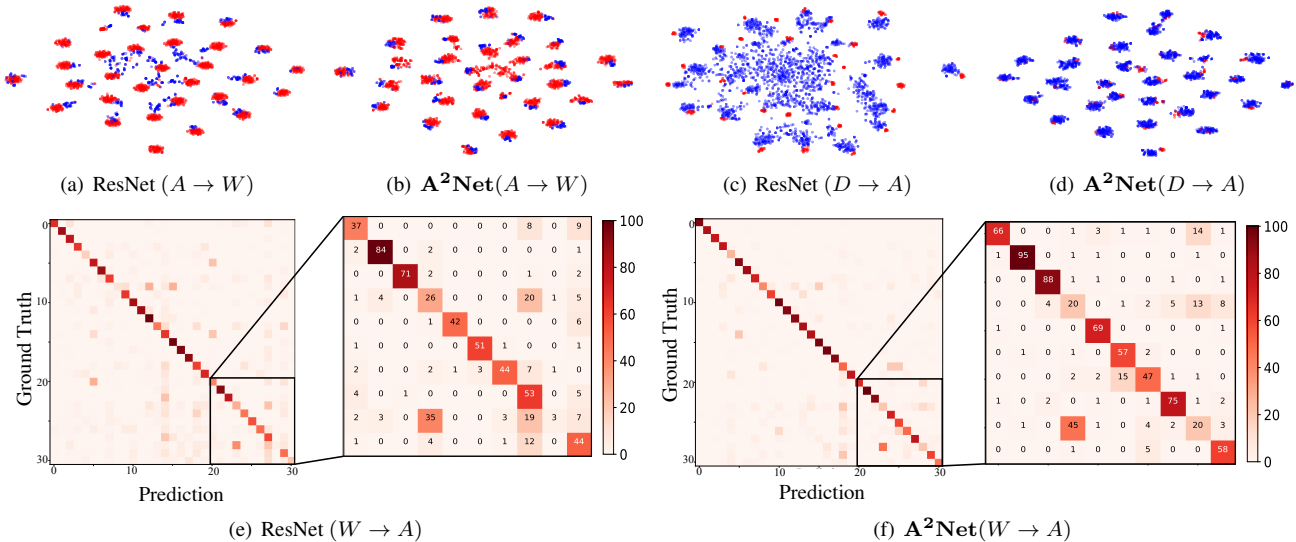(e) ResNet $(W \rightarrow A)$ (f) $\mathbf{A^2Net}(W \rightarrow A)$

Figure 3. Result of Feature Visualization and Confusion Matrix. (a)-(d) show high-level source (red) and target (blue) features generated by source-only model (Resnet-50) and our $\mathbf{A^2Net}$. Note that we only exploit source data to draw the t-SNE without any use of it during adaptation stage. (e) and (f) are the confusion matrices, comparing the ground-truth and the category prediction from ResNet and our model, respectively.
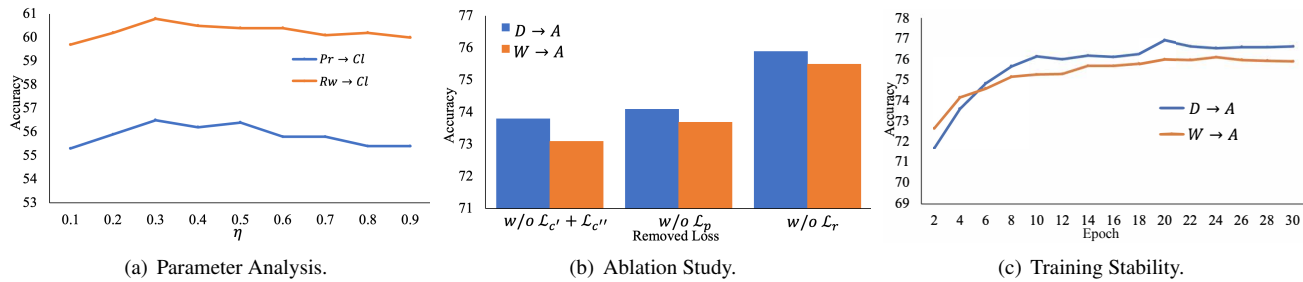

(a) Parameter Analysis.   (b) Ablation Study.   (c) Training Stability.

Figure 4. (a) Parameter Analysis records the object recognition accuracy with the varying η. (b) Ablation Study shows the influence of removing each constraint on the performance of our model. (c) Training Stability reports the object recognition ability of target classifier as the increasing number of epoch.

performance of the frozen source classifier. Similarly, removing the contrastive category matching also results in the performance degradation since this module mainly exploits the existed knowledge of source model to explore the relation of any two target samples and controls the compactness of each target class subspace by using contrastive loss over all positive pairs. In terms of the rotation design, it actually makes a small contribution to the improvement of performance by learning additional semantics from target images in self-supervised manner. However, we still promote the adaptation ability of model via the adjustment of parameter η balancing the rotation constraint and others. For instance, Fig. 4 (a) reports the relation between the varying η and target classification accuracy. These two tasks of Office-Home both achieve the highest performance with $\eta = 0.3$. Finally, considering the adversarial game between feature generator and dual-classifier , we show the change of object recognition accuracy as the increasing of epoch in Fig. 4 (c). With the adversarial training manner, the target classifier gradually improves its classification ability in a stable rhythm.

5. Conclusions

Unsupervised Domain Adaptation (UDA) assumes the well-annotated source domain and unlabeled target images are both available for the model training. However, many practical applications only access the well-trained source model instead of source data during adaptation stage, which is defined as source-free domain adaptation. To overcome the novel scenario, this paper proposes Adaptive Adversarial Network ($\mathbf{A^2Net}$) including three operations. First, $\mathbf{A^2Net}$ develops a soft-adversarial mechanism to learn a flexible target classifier promoting the recognition of samples which the frozen source classifier difficultly identifies. Second, it explores the contrastive loss over all positive paired target samples to intensify the compactness of each category subspace. Finally, the self-supervised rotation is adopted to learn additional semantics from target images to learn more discriminative features. Moreover, experiments of three popular benchmarks illustrate our method effectively achieves domain adaptation without source data.

References

[1] 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, 5

[2] 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. 1

[3] Yiming Chen, Shiji Song, Shuang Li, and Cheng Wu. A graph embedding framework for maximum mean discrepancy-based domain adaptation algorithms. *IEEE Transactions on Image Processing*, 29:199–213, 2019. 1

[4] Shuhao Cui, Shuhui Wang, Junbao Zhuo, Liang Li, Qingming Huang, and Qi Tian. Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3941–3950, 2020. 6, 7

[5] 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. 2

[6] 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] Jiahua Dong, Yang Cong, Gan Sun, Yuyang Liu, and Xiaowei Xu. Cscl: Critical semantic-consistent learning for unsupervised domain adaptation. In *European Conference on Computer Vision*, pages 745–762. Springer, 2020. 1

[8] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016. 6, 7

[9] 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. 1

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 6, 7

[11] Ying Jin, Ximei Wang, Mingsheng Long, and Jianmin Wang. Minimum class confusion for versatile domain adaptation. In *European Conference on Computer Vision*, pages 464–480. Springer, 2020. 6, 7

[12] Taotao Jing and Zhengming Ding. Adversarial dual distinct classifiers for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 605–614, 2021. 2

[13] Taotao Jing, Haifeng Xia, and Zhengming Ding. Adaptively-accumulated knowledge transfer for partial domain adaptation. In *Proceedings of the 28th ACM International Conference on Multimedia*, pages 1606–1614, 2020. 2

[14] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4893–4902, 2019. 1, 2

[15] Youngeun Kim, Sungeun Hong, Donghyeon Cho, Hyoungseob Park, and Priyadarshini Panda. Domain adaptation without source data. *arXiv preprint arXiv:2007.01524*, 2020. 6, 7

[16] Atsutoshi Kumagai and Tomoharu Iwata. Unsupervised domain adaptation by matching distributions based on the maximum mean discrepancy via unilateral transformations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 4106–4113, 2019. 1

[17] Vinod K Kurmi, Venkatesh K Subramanian, and Vinay P Namboodiri. Domain impression: A source data free domain adaptation method. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 615–625, 2021. 3, 6

[18] Mengxue Li, Yi-Ming Zhai, You-Wei Luo, Peng-Fei Ge, and Chuan-Xian Ren. Enhanced transport distance for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13936–13944, 2020. 2

[19] Rui Li, Qianfen Jiao, Wenming Cao, Hau-San Wong, and Si Wu. Model adaptation: Unsupervised domain adaptation without source data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9641–9650, 2020. 2, 3, 6

[20] Justin Liang, Namdar Homayounfar, Wei-Chiu Ma, Yuwen Xiong, Rui Hu, and Raquel Urtasun. Polytransform: Deep polygon transformer for instance segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9131–9140, 2020. 1

[21] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International Conference on Machine Learning*, pages 6028–6039. PMLR, 2020. 2, 3, 5, 6, 7

[22] 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. 5

[23] Hongfu Liu, Ming Shao, Zhengming Ding, and Yun Fu. Structure-preserved unsupervised domain adaptation. *IEEE Transactions on Knowledge and Data Engineering*, 2018. 3

[24] Ziming Liu, Guangyu Gao, Lin Sun, and Li Fang. Ipgnet: Image pyramid guidance network for small object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 1026–1027, 2020. 1

[25] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. *arXiv preprint arXiv:1705.10667*, 2017. 6, 7

[26] 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. 1, 2

[27] Zhongyi Pei, Zhangjie Cao, Mingsheng Long, and Jianmin Wang. Multi-adversarial domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018. 1

[28] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1406–1415, 2019. 1

[29] Xingchao Peng, Ben Usman, Neela Kaushik, Dequan Wang, Judy Hoffman, and Kate Saenko. Visda: A synthetic-to-real benchmark for visual domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 2021–2026, 2018. 5

[30] Kate Saenko, Brian Kulis, Mario Fritz, and Trevor Darrell. Adapting visual category models to new domains. In *European conference on computer vision*, pages 213–226. Springer, 2010. 5

[31] 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

[32] Hui Tang, Ke Chen, and Kui Jia. Unsupervised domain adaptation via structurally regularized deep clustering. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8725–8735, 2020. 6, 7

[33] 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. 1, 2

[34] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5018–5027, 2017. 3, 5

[35] Haifeng Xia and Zhengming Ding. Hgnet: Hybrid generative network for zero-shot domain adaptation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16*, pages 55–70. Springer, 2020. 2

[36] Haifeng Xia and Zhengming Ding. Structure preserving generative cross-domain learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4364–4373, 2020. 1

[37] Cai Xu, Ziyu Guan, Wei Zhao, Hongchang Wu, Yunfei Niu, and Beilei Ling. Adversarial incomplete multi-view clustering. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 3933–3939. AAAI Press, 2019. 7

[38] 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/CVF International Conference on Computer Vision*, pages 1426–1435, 2019. 2, 6, 7

[39] Renjun Xu, Pelen Liu, Liyan Wang, Chao Chen, and Jindong Wang. Reliable weighted optimal transport for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4394–4403, 2020. 1

[40] Zhe Xu, Shaoli Huang, Ya Zhang, and Dacheng Tao. Webly-supervised fine-grained visual categorization via deep domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 40(5):1100–1113, 2018. 3

[41] 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. 1

[42] Guanglei Yang, Haifeng Xia, Mingli Ding, and Zhengming Ding. Bi-directional generation for unsupervised domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 6615–6622, 2020. 1

[43] Hao-Wei Yeh, Baoyao Yang, Pong C Yuen, and Tatsuya Harada. Sofa: Source-data-free feature alignment for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 474–483, 2021. 2, 6, 7

[44] Manli Zhang, Jianhong Zhang, Zhiwu Lu, Tao Xiang, Mingyu Ding, and Songfang Huang. Iept: Instance-level and episode-level pre-text tasks for few-shot learning. 5

[45] Weichen Zhang, Wanli Ouyang, Wen Li, and Dong Xu. Collaborative and adversarial network for unsupervised domain adaptation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 3801–3809, 2018. 3

[46] Yabin Zhang, Hui Tang, Kui Jia, and Mingkui Tan. Domain-symmetric networks for adversarial domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5031–5040, 2019. 2

[47] Handong Zhao, Zhengming Ding, and Yun Fu. Multi-view clustering via deep matrix factorization. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017. 1

[48] Handong Zhao, Hongfu Liu, and Yun Fu. Incomplete multi-modal visual data grouping. In *IJCAI*, pages 2392–2398, 2016. 1

[49] Kaiyang Zhou, Yongxin Yang, Timothy Hospedales, and Tao Xiang. Learning to generate novel domains for domain generalization. In *European Conference on Computer Vision*, pages 561–578. Springer, 2020. 1