Concurrent Subsidiary Supervision for Unsupervised Source-Free Domain Adaptation

Jogendra Nath Kundu[1]*, Suvaansh Bhambri[1]*, Akshay Kulkarni[1]*, Hiran Sarkar[1], Varun Jampani[2], and R. Venkatesh Babu[1]

[1] Indian Institute of Science
[2] Google Research

Abstract. The prime challenge in unsupervised domain adaptation (DA) is to mitigate the domain shift between the source and target domains. Prior DA works show that pretext tasks could be used to mitigate this domain shift by learning domain invariant representations. However, in practice, we find that most existing pretext tasks are ineffective against other established techniques. Thus, we theoretically analyze how and when a subsidiary pretext task could be leveraged to assist the goal task of a given DA problem and develop objective subsidiary task suitability criteria. Based on this criteria, we devise a novel process of sticker intervention and cast sticker classification as a supervised subsidiary DA problem concurrent to the goal task unsupervised DA. Our approach not only improves goal task adaptation performance, but also facilitates privacy-oriented source-free DA i.e. without concurrent source-target access. Experiments on the standard Office-31, Office-Home, DomainNet, and VisDA benchmarks demonstrate our superiority for both single-source and multi-source source-free DA. Our approach also complements existing non-source-free works, achieving leading performance.

1 Introduction

The prevalent trend in supervised deep learning systems is to assume that training and testing data follow the same distribution. However, such models often fail [6] when deployed in a new environment (target domain) due to the discrepancy in the training (source domain) and target distributions. A standard approach to deal with this problem of *domain shift* is Unsupervised Domain Adaptation (DA) [10,30], which aims to minimize the domain discrepancy [3] between source and target. The prime challenge in DA is to facilitate the effective utilization of the unlabeled samples while adapting to the target domain.

Drawing motivation from self-supervised pretext task literature [35,13], recent DA works [5,31] have adopted subsidiary tasks as side-objectives to improve the adaptation performance. The intuition is that subsidiary task objectives enforce learning of domain-generic representations, leading to improved domain alignment [51] and consequently, better feature clustering for unlabeled target

* Equal contribution | Webpage: https://sites.google.com/view/sticker-sfda

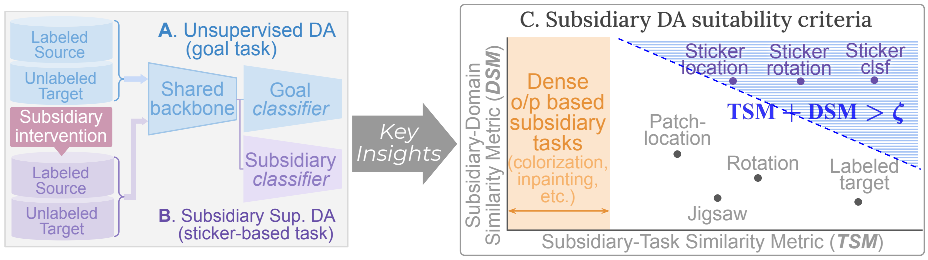

Fig. 1. We tackle **A.** unsupervised goal task DA by introducing **B.** a concurrent subsidiary supervised DA. **C.** Our theoretical insights reveal that subsidiary tasks having both higher TSM (X-axis) and DSM (Y-axis) are most suitable for concurrent goal-subsidiary adaptation (*i.e.* the shaded blue area). The proposed sticker-based tasks better suit concurrent goal-subsidiary DA among other self-supervised pretext tasks.

[31]. We aim to design a similar framework but, contrary to prior works, we adopt a novel perspective of subsidiary supervised DA for the subsidiary task concurrent to unsupervised goal task DA. Specifically, the framework involves a shared backbone with a goal classifier and a subsidiary classifier (Fig. 1A, B).

To better understand how subsidiary supervised DA objectives support goal task DA, we intend to theoretically analyze the proposed framework. While several subsidiary tasks are available in the literature, there has been little attention on identifying the desirable properties of a subsidiary task that would better aid the unsupervised DA. A recent self-supervised work [55] studied the effectiveness of pretraining with existing subsidiary tasks [35,13] on different downstream supervised settings such as fine-grained or medical image classification [37,56]. We argue that our intended theoretical analysis is necessary to understand the same for DA settings as DA presents a different set of challenges compared to downstream supervised learning paradigms.

Thus, we attempt to answer two interconnected questions,

1. *How does subsidiary supervised DA help goal task unsupervised DA?*

2. *What kind of subsidiary tasks better suit concurrent goal-subsidiary DA?*

For the first question, we uncover theoretical insights based on generalization bounds in DA [3,64]. These bounds define distribution shift or domain discrepancy between source and target as the worst discrepancy for a given hypothesis space. We analyze the effect of adding the subsidiary supervised DA problem on the hypothesis space of the shared backbone. Based on this, we find that a higher domain similarity between goal and subsidiary task samples leads to a lower domain discrepancy. This leads to better adaptation for concurrent goal-subsidiary DA w.r.t. naive goal DA. Further, we observe that a higher goal-subsidiary task similarity aids effective learning of both tasks with the shared backbone, which is crucial for subsidiary DA to positively impact the goal DA.

For the second question, we first devise a subsidiary-domain similarity metric (DSM) and a subsidiary-task similarity metric (TSM) to measure the domain similarity and task similarity between any subsidiary task with a given goal task.

Based on our theoretical insights, we propose a subsidiary task suitability criteria using both DSM and TSM to identify *DA-assistive* subsidiary tasks. With this criteria, we evaluate the commonly used subsidiary tasks from the pretext task literature like rotation prediction [31], patch location [51], and jigsaw permutation prediction [5] in Fig. 1C. We observe that these existing tasks have significantly low DSM. On the other hand, dense output based tasks like colorization [22] or inpainting [38] severely lack in TSM as goal task is classification-based. Understanding these limitations, we devise a sticker-intervention that facilitates domain preservation (high DSM) and propose a range of sticker-based subsidiary tasks (Fig. 2). For general shape-based goal tasks, it turns out that sticker classification task has the best TSM among other sticker-based tasks. This yields higher adaptation performance thereby validating the proposed criteria.

To evaluate our theoretical insights and the proposed concurrent subsidiary DA, we particularly focus on source-free DA regime [23,21,17]. In this, the source and target data are not concurrently accessible while model sharing is permitted. While this challenging setting holds immense practical value by working within the data privacy regulations, we choose source-free DA as it can prominently highlight our advantages. The well-developed discrepancy minimization techniques, tailored to general DA scenarios, guide the adaptation more significantly than our proposed approach but cannot be used for source-free DA. Further, existing source-free works [26] rely heavily on pseudo-label based self-training on target data. Our proposed subsidiary supervised adaptation implicitly regularizes target-side self-training, leading to improved adaptation.

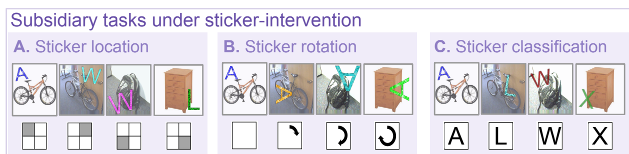

Fig. 2. Sticker intervention involves mixup of input with a masked sticker. We devise the following sticker-based tasks; **A.** locating the quadrant of the sticker, **B.** predicting sticker rotation, **C.** classifying sticker category.

To summarize, our main contributions are:

- We introduce concurrent subsidiary supervised DA, for a subsidiary task, that not only improves unsupervised goal task DA but also facilitates source-free adaptation. We provide theoretical insights to analyze the impact of subsidiary DA on the domain discrepancy, and hence, the goal task DA.
- Based on our insights, we devise a subsidiary DA suitability criteria to identify *DA-assistive* subsidiary tasks that better aid the unsupervised goal task DA. We also propose novel sticker intervention based subsidiary tasks that demonstrate the efficacy of the criteria.
- Our proposed approach achieves state-of-the-art performance on source-free single-source DA (SSDA) as well as source-free multi-source DA (MSDA) for image classification. The proposed approach also complements existing non-source-free works, achieving leading performance.

2 Related Work

Pretext tasks in self-supervised learning. Pretext tasks are used to learn deep feature representations from unlabeled data, in a self-supervised manner, for downstream tasks. There are several pretext tasks such as image inpainting [38], colorization [62,22,63], spatial context prediction [7], contrastive predictive coding [36], image rotation [13], and jigsaw puzzle solving [35]. Pretext tasks are commonly used for pre-training on unlabeled data followed by finetuning on labeled data. Conversely, we perform supervised DA for the pretext-like task along with the unsupervised goal task DA, resulting in a representation that aligns the domains while maintaining the goal task performance.

Source-free DA. Recently, several methods have investigated source-free DA. USFDA [19] and FS [20] investigate universal DA [61] and open-set DA [46], in a source-free setting by synthesizing training samples to make the decision boundaries compact. SHOT [26,27], NRC [59] maximize mutual information and propose pseudo-labeling, using global structure to match target features to that of a fixed source classifier. To provide adaptation supervision, 3C-GAN [23] generates labeled target-style images from a GAN. Finally, SFDA [28], UR [49], and GtA [18] are semantic segmentation specific source-free DA techniques.

Pretext task based DA. Several DA works have demonstrated the efficacy of learning meaningful representations using pretext tasks. Early works [11,12] used reconstruction as a pretext task to extract domain-invariant features. [4] captured both domain-specific and shared features by separating the feature space into domain-private and domain-shared spaces. [5] used jigsaw puzzles as a side-objective to tackle domain generalization. [51] proposed that adaptation can be accomplished by learning many self-supervision tasks at the same time. [16] suggested a cross-domain SSL strategy for adaptation with minimal source labels based on instance discrimination [57]. [15] recommended employing SSL pretext tasks like rotation prediction and patch placement prediction. [45] solved the challenge of universal domain adaptation by unsupervised clustering. [43] employed easy labels for synthetic images, such as the surface normal, depth, and instance contour, to train a network. [9] employed SSL pretext tasks like rotation prediction as part of their domain generalization technique.

3 Approach

In this section, we introduce required preliminaries (Sec. 3.1), followed by theoretical insights (Sec. 3.2) that motivate our training algorithm design (Sec. 3.4).

3.1 Preliminaries

3.1.1 Goal task unsupervised DA. For closed set DA problem, consider a labeled source dataset $\boldsymbol{\mathcal{D}_s = \{(x_s, y_s) : x_s \in \mathcal{X}, y_s \in \mathcal{C}_g\}}$ where $\mathcal{X}$ is the input space and $\mathcal{C}_g$ denotes the label set for the goal task. x_s is drawn from the marginal distribution p_s. Let $\boldsymbol{\mathcal{D}_t = \{x_t : x_t \in \mathcal{X}\}}$ be an unlabeled target dataset with

$x_t \sim p_t$. The goal is to assign labels for each target image x_t. The usual approach [10,50,29] is to use a backbone feature extractor $h : \mathcal{X} \to \mathcal{Z}$ followed by a goal classifier $f_g : \mathcal{Z} \to \mathcal{C}_g$ (see Fig. 3A). The expected source risk with h and an optimal labeling function $f_S : \mathcal{X} \to \mathcal{C}_g$, is $\epsilon_s(h) = \mathbb{E}_{x \sim p_s}[\mathbb{1}(f_g \circ h(x) \neq f_S(x))]$, where (.) is an indicator function. Similarly, $\epsilon_t(h)$ is the target risk with optimal labeling function $f_T : \mathcal{X} \to \mathcal{C}_g$. We restate the theoretical upper bound on target risk from [64]. For backbone hypothesis $h \in \mathcal{H}$ with $\mathcal{H}$ being the hypothesis space and a domain classifier $f_d : \mathcal{Z} \to \{0, 1\}$ (0 for source, 1 for target),

$$\begin{aligned} \epsilon_t(h) &\leq \epsilon_s(h) + d_{\mathcal{H}}(p_s, p_t) + \lambda_g; \\ \text{where, } \lambda_g &= \min\left\{\mathbb{E}_{p_s}[\mathbb{1}(f_S(x) \neq f_T(x))], \mathbb{E}_{p_t}[\mathbb{1}(f_S(x) \neq f_T(x))]\right\} \\ \text{and, } d_{\mathcal{H}}(p_s, p_t) &= \sup_{h \in \mathcal{H}} \left| \mathbb{E}_{x \sim p_s}[\mathbb{1}(f_d \circ h(x) = 1)] - \mathbb{E}_{x \sim p_t}[\mathbb{1}(f_d \circ h(x) = 1)] \right| \end{aligned} \tag{1}$$

Here, $d_{\mathcal{H}}$ is the $\mathcal{H}$-divergence [3] that indicates the distribution shift or worst-case domain discrepancy between the two domains. λ_g is a constant that represents the optimal cross-domain error of the labeling functions. Thus, the target risk $\epsilon_t(h)$ is upper bounded by these two terms along with the source risk $\epsilon_s(h)$.

3.1.2 Subsidiary supervised DA. Next, we introduce a subsidiary supervised DA problem concurrent to the goal task unsupervised DA. To this end, we aim to devise a subsidiary classification task with a new label set $\mathcal{C}_n$. The label-set specific attributes are inflicted on $x \in \mathcal{X}$ via an intervention, to form supervised pairs. These pairs form labeled source, $(\boldsymbol{x_{s,n}}, \boldsymbol{y_n}) \in \boldsymbol{\mathcal{D}_{s,n}}$ and labeled target, $(\boldsymbol{x_{t,n}}, \boldsymbol{y_n}) \in \boldsymbol{\mathcal{D}_{t,n}}$ datasets. Here, the inputs $x_{s,n}$ and $x_{t,n}$ are drawn from marginal distributions $p_{s,n}$ and $p_{t,n}$ respectively. We also define the optimal labeling functions for source and target subsidiary task as $f_{S,n} : \mathcal{X} \to \mathcal{C}_n$ and $f_{T,n} : \mathcal{X} \to \mathcal{C}_n$. Next, the prediction mapping involves the shared goal-task backbone h followed by a subsidiary classifier $f_n : \mathcal{Z} \to \mathcal{C}_n$ (see Fig. 3A). Here, the source-subsidiary task error is $\epsilon_{s,n}(h) = \mathbb{E}_{x \sim p_{s,n}}[\mathbb{1}(f_n \circ h(x) \neq f_{S,n}(x))]$. Similarly, $\epsilon_{t,n}(h)$ for target and λ_n defined as in Eq. 1. Thus, generalization bounds for subsidiary DA with the same $\mathcal{H}$ is stated as,

$$\epsilon_{t,n}(h) \leq \epsilon_{s,n}(h) + d_{\mathcal{H}}(p_{s,n}, p_{t,n}) + \lambda_n \tag{2}$$

3.1.3 Metrics. We introduce two metrics that form the basis of our insights.
a) Subsidiary-Domain Similarity Metric (DSM), $\boldsymbol{\gamma_{DSM}(.,.)}$. DSM measures the similarity between two domains as the inverse of the standard $\mathcal{A}$-distance [3]. $\mathcal{A}$-distance can be thought of as a proxy [10] for $\mathcal{H}$-divergence.
b) Subsidiary-Task Similarity Metric (TSM), $\boldsymbol{\gamma_{TSM}(.,.)}$. TSM measures the task similarity of a subsidiary task w.r.t. the goal task. TSM is computed using the standard linear evaluation protocol [47] borrowed from transfer learning and self-supervised literature. It is the performance of a subsidiary-task linear

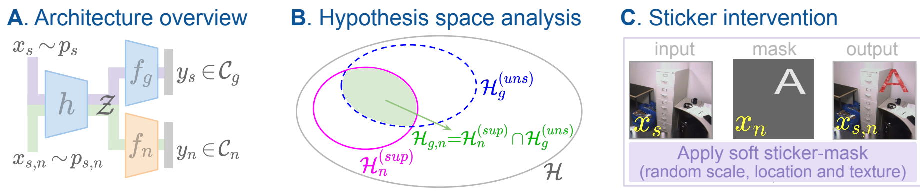

Fig. 3. A. Our method uses a shared backbone h with goal classifier f_g and subsidiary classifier f_n. **B.** Hypothesis space analysis for only goal DA, only subsidiary DA and concurrent goal-subsidiary DA (Sec. 3.2.1). **C.** Sticker intervention.

classifier attached to a goal-task pretrained backbone feature extractor $h_{s,g}$. Intuitively, it indicates the extent of compatibility between the two tasks.

For a dataset pair of source-goal and source-subsidiary, *i.e.* $(\mathcal{D}_s, \mathcal{D}_{s,n})$;

$$\gamma_{DSM}(\mathcal{D}_s, \mathcal{D}_{s,n}) = 1 - \frac{1}{2} d_{\mathcal{A}}(\mathcal{D}_s, \mathcal{D}_{s,n}); \quad \gamma_{TSM}(\mathcal{D}_s, \mathcal{D}_{s,n}) = 1 - \min_{f_n} \hat{\epsilon}_{s,n}(h_{s,g}) \quad (3)$$

Here, $d_{\mathcal{A}}(.,.)$ denotes $\mathcal{A}$-distance and $\hat{\epsilon}_{s,n}(.)$ denotes empirical error for subsidiary task on source data. Note that $0 \leq \hat{\epsilon}_{s,n}(h_{s,g}) \leq 1$ while $0 \leq d_{\mathcal{A}}(\mathcal{D}_1, \mathcal{D}_2) \leq 2$.

3.2 Theoretical insights

We analyze the impact of solving subsidiary supervised DA on the goal task unsupervised DA. We first consider the combined bounds (combining Eq. 1, 2),

$$\epsilon_t(h) + \epsilon_{t,n}(h) \leq \epsilon_s(h) + \epsilon_{s,n}(h) + d_{\mathcal{H}}(p_s, p_t) + d_{\mathcal{H}}(p_{s,n}, p_{t,n}) + \lambda_g + \lambda_n \quad (4)$$

Among the six terms on the right side, the two λ terms are constants as they do not involve the hypothesis h or hypothesis space $\mathcal{H}$. We analyze the source error duet, $\epsilon_s(h) + \epsilon_{s,n}(h)$, and the domain discrepancy duet $d_{\mathcal{H}}(p_s, p_t) + d_{\mathcal{H}}(p_{s,n}, p_{t,n})$.

3.2.1 Analyzing the domain discrepancy duet (Fig. 3B). We analyze w.r.t. the domain discrepancy duet considering three configurations:

a) While performing only unsupervised goal task DA, the backbone optimization would operate on a limited hypothesis space $\mathcal{H}_g^{(uns)} \subset \mathcal{H}$ where $\mathcal{H}_g^{(uns)} = \{h \in \mathcal{H} : |\epsilon_t(h) - \epsilon_s(h)| \leq \zeta_g^{(uns)}\}$. Here, $\zeta_g^{(uns)}$ is a threshold on the source-target error gap.

b) While performing supervised adaptation only for subsidiary domain adaptation, the optimization would operate on a limited hypothesis space $\mathcal{H}_n^{(sup)} \subset \mathcal{H}$ *i.e.*, $\mathcal{H}_n^{(sup)} = \{h \in \mathcal{H} : |\epsilon_{t,n}(h) - \epsilon_{s,n}(h)| \leq \zeta_n^{(sup)}\}$. Here, $\zeta_n^{(sup)}$ is a threshold on the subsidiary-task source-target error gap.

c) While concurrently performing **a)** unsupervised goal task DA and **b)** subsidiary supervised DA (*i.e.* the proposed approach), the optimization would operate on a limited hypothesis space $\mathcal{H}_{g,n} \subset \mathcal{H}$. Specifically, $\mathcal{H}_{g,n} = \mathcal{H}_n^{(sup)} \cap \mathcal{H}_g^{(uns)}$. This is because the backbone is shared between the two DA tasks and hence, would be limited to the intersection space.

Different configurations lead to different $\mathcal{H}$-spaces and consequently, different $\mathcal{H}$-divergences. Comparing the $\mathcal{H}$-divergences leads us to the following insight.

Insight 1. ($\mathcal{H}$-divergence in concurrent goal DA and subsidiary DA) *The backbone hypothesis space for concurrent unsupervised goal DA and subsidiary supervised DA, i.e.* $\mathcal{H}_{g,n} = \mathcal{H}_n^{(sup)} \cap \mathcal{H}_g^{(uns)}$ *will yield a lower* $\mathcal{H}$*-divergence than* $\mathcal{H}_g^{(uns)}$ *(hypothesis space for only unsupervised goal task DA), i.e.*

$$d_{\mathcal{H}_{g,n}}(p_s, p_t) \leq d_{\mathcal{H}_g^{(uns)}}(p_s, p_t) \text{ and } d_{\mathcal{H}_{g,n}}(p_{s,n}, p_{t,n}) \leq d_{\mathcal{H}_g^{(uns)}}(p_{s,n}, p_{t,n}) \quad (5)$$

Remarks. In Eq. 1, $d_{\mathcal{H}}(p_s, p_t)$ is the supremum over the hypothesis space $\mathcal{H}$ *i.e.* a worst-case measure. Since $\mathcal{H}_{g,n} \subset \mathcal{H}_g^{(uns)}$, $\mathcal{H}_{g,n}$ would have a lower $\mathcal{H}$-divergence as the worst-case hypothesis of $\mathcal{H}_g^{(uns)}$ may be absent in the subset $\mathcal{H}_{g,n}$. This applies to both pairs, (p_s, p_t) and $(p_{s,n}, p_{t,n})$. While a lower $\mathcal{H}$-divergence duet leads to improved goal DA, the equality may hold when the worst hypothesis of $\mathcal{H}_g^{(uns)}$ remains in $\mathcal{H}_{g,n}$. In such a case, concurrent DA would perform the same as naive goal DA. To this end, we put forward the following insight.

Insight 2. (When is concurrent DA strictly better than naive DA?) *A subsidiary task supports the strict inequality* $d_{\mathcal{H}_{g,n}}(p_s, p_t) < d_{\mathcal{H}_g^{(uns)}}(p_s, p_t)$ *if with at least* $(1-\delta)$ *probability, the subsidiary-domain similarity* $\gamma_{DSM}(\mathcal{D}_s, \mathcal{D}_{s,n})$ *exceeds a threshold* ζ_d *by no less than* ξ; $\mathbb{P}[\boldsymbol{\gamma_{DSM}(\mathcal{D}_s, \mathcal{D}_{s,n}) \geq \zeta_d - \xi}] \boldsymbol{\geq 1 - \delta}$.

Remarks. In other words, the strict inequalities in Eq. 5 would hold if the DSM $\gamma_{DSM}(.,.)$ exceeds a threshold ζ_d. The supports for this insight are twofold. First, a subsidiary task may heavily alter domain information [32], *e.g.* jigsaw shuffling [5]. Then, the backbone will be updated using out-of-domain samples which is undesirable as such samples are unlikely for inference. This will be avoided if Insight 2 is satisfied. Second, if DSM is high, we can approximate $p_s \approx p_{s,n}$ and $p_t \approx p_{t,n}$. Thus, more samples from subsidiary task data will be available for training the backbone to be domain-invariant (as subsidiary task uses samples from both the domains) *i.e.* reducing $d_{\mathcal{H}}$ against the same in naive goal DA.

3.2.2 Analyzing the source error duet.

Now we analyze w.r.t. the source error duet of Eq. 4. While the $\mathcal{H}$-divergence is lower for concurrent goal task DA and subsidiary supervised DA, a logical concern is that simultaneous minimization of errors, *i.e.* $\epsilon_s(h) + \epsilon_{s,n}(h)$, for both tasks may be difficult with the shared backbone h. Further, it may happen that simultaneous training for both tasks in target domain may hamper the goal task performance as it is unsupervised. In such cases, the subsidiary task would be ill-equipped to assist the goal task adaptation. To avoid these, we propose another empirical criterion as follows.

Insight 3. (Goal and subsidiary task similarity for concurrent DA) *Higher goal-subsidiary task similarity (TSM) aids effective minimization of both task errors with the shared backbone, which is crucial for subsidiary supervised DA to positively affect the goal task DA. The criterion is* $\boldsymbol{\gamma_{TSM}(\mathcal{D}_s, \mathcal{D}_{s,n}) > \zeta_n}$.
Remarks. Here, ζ_n is a threshold. The TSM γ_{TSM} indicates the compatibility of goal task features to support the subsidiary task. Intuitively, a higher TSM implies more overlap in the discriminative features of the two tasks, which would allow better simultaneous minimization of both task errors.

Based on Insight 1, concurrent subsidiary supervised DA and goal task DA yields a lower domain discrepancy. Further, based on Insight 2, a subsidiary task can be selected such that effective minimization of both source errors is possible simultaneously. Thus, using Eq. 1, we can infer that $\sup_{h\in\mathcal{H}_{g,n}} \epsilon_t(h) \leq \sup_{h\in\mathcal{H}_g^{(uns)}} \epsilon_t(h)$ *i.e.* a lower target error upper bound for our approach w.r.t. naive goal task DA. Now, we summarize the criteria (Insight 2, 3).
Definition 1. (Subsidiary DA suitability criteria) *A subsidiary task is termed DA-assistive i.e. suitable for subsidiary supervised DA if the sum of DSM* γ_{DSM} *and TSM* γ_{TSM} *exceeds a threshold* ζ,

$$\gamma_{DSM}(\mathcal{D}_s, \mathcal{D}_{s,n}) + \gamma_{TSM}(\mathcal{D}_s, \mathcal{D}_{s,n}) > \zeta \tag{6}$$

Remarks. In other words, a subsidiary task which is domain-preserving and has high task similarity w.r.t. the goal task is *DA-assistive i.e.* suitable for subsidiary supervised DA to aid the goal task DA. We employ this criteria empirically for a diverse set of subsidiary tasks (shown in Fig. 1C). Next, we describe the motivation for our proposed sticker intervention and corresponding subsidiary tasks as well as training algorithms tailored for source-free DA.

3.3 Sticker intervention based subsidiary task design

While one may consider pretext tasks from the self-supervised learning literature as candidates for subsidiary DA, almost all such tasks fail to satisfy subsidiary DA suitability criteria in Eq. 6. For instance, dense output based tasks such as colorization [62,22], inpainting [38], *etc.* exhibit markedly low task similarity (TSM) against the non-dense goal tasks. Further, the input intervention for certain pretext tasks such as jigsaw [5], patch-location[51], rotation [31,15], significantly alter the domain information leading to low domain similarity (DSM).
Insight 4. (Sticker-intervention based tasks well suit subsidiary DA) *Sticker intervention is the process of pasting a sticker* x_n *(i.e., a symbol with random texture and scale) on a given image sample* $x_s \in \mathcal{D}_s$ *to obtain a stickered sample, i.e.* $x_{s,n} = \mathcal{T}(x_s, x_n) \in \mathcal{D}_{s,n}$. *Following this, the subsidiary task could be defined as the classification of some sticker attribute (e.g. shape, location, or orientation). Such a formalization provides effective control to maximize* $\gamma_{DSM}(\mathcal{D}_s, \mathcal{D}_{s,n})$ *and* $\gamma_{TSM}(\mathcal{D}_s, \mathcal{D}_{s,n})$, *in line with our suitability criteria.*
Remarks. The sticker intervention (Fig. 3C) facilitates domain preservation while simultaneously supporting a range of subsidiary tasks. Since the proposed sticker intervention alters only a local area of the sample, the original content is

not suppressed which in turn preserves the domain information, implying high DSM. Following this, one can ablate over a range of sticker-based tasks in order to select a suitable subsidiary task based on the given goal task. Below, we discuss some possible subsidiary tasks under the sticker intervention.

a) Sticker location (Fig. 2A). We draw motivation from patch-location [51], where the task is to classify the quadrant to which a patch-input belongs. With sticker intervened images, the task is to classify the quadrant with the sticker. Our use of whole images as input is more domain-preserving than patch-input.

b) Sticker rotation (Fig. 2B). Motivated by the image rotation task [31], we propose sticker rotation task where the rotation of the sticker has to be classified (0°, 90°, 180°and 270°rotations possible). Note that our sticker rotation does not affect the domain information while rotating the entire image does.

c) Sticker classification (Fig. 2C). While the discriminative features in the previous two tasks were location and rotation, we propose sticker classification task with primary discriminative features as shape. In other words, the task is to classify the sticker shape (*i.e.* the symbol) given a stickered sample.

3.4 Training algorithm design under source-free constraints

For the standard DA setting with concurrent access to source and target data [10,50], the subsidiary supervised DA can be implemented simply by optimizing the subsidiary classification loss simultaneously for source and target. This would yield a lower domain discrepancy as discussed in Sec. 3.2. However, in the more practical source-free setting [23,19] where concurrent source-target access is prohibited, this simple approach would not be possible. We believe the improvements will be prominent in source-free DA based on the following insight:

Insight 5. (Subsidiary DA better suits challenging source-free DA). *Existing source-free DA works heavily rely on pseudo-label or clustering based self-training on unlabeled target with no obvious alternative. The proposed subsidiary supervised adaptation helps to implicitly regularize the target-side self-training, leading to improved adaptation performance. The subsidiary DA not only aids goal DA as a result of high DSM but also preserves the goal task inductive bias as a result of high TSM, while adhering to the source-free constraints.*

Remarks. The source-free setting presents new challenges which highlight the advantages of our proposed method more prominently. This is because, the performance in non-source-free DA is strongly influenced by well-developed discrepancy minimization techniques. However, these techniques cannot be leveraged in a source-free setting due to their requirement of concurrent source-target data access. Thus, we primarily operate in the source-free regime to evaluate our theoretical insights and the proposed concurrent subsidiary supervised DA problem.

We perform the training in three steps. First two steps involve pre-training of goal task and subsidiary task respectively with source data. The final step involves adapting both tasks to target domain. For clarity, we first summarize available and intervened datasets required for training and their notations.

Datasets. The goal task source data is denoted by $(x_s, y_s) \in \mathcal{D}_s$ while the corresponding unlabeled target is denoted by $x_t \in \mathcal{D}_t$. The intervened stickered-

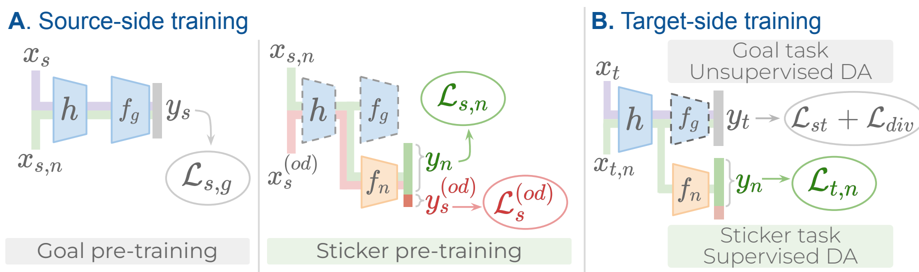

Fig. 4. **A.** Source-side training involves goal pre-training (Sec. 3.4.1) and sticker pre-training (Sec. 3.4.2). **B.** Target-side training involves concurrent goal-task unsupervised DA and sticker-task supervised DA (Sec. 3.4.3).

source data, coupled with both goal and sticker task labels, is denoted by $(x_{s,n}, y_s, y_n) \in \mathcal{D}_{s,n}$. The corresponding stickered-target data, with only subsidiary sticker task labels, is denoted by $(x_{t,n}, y_n) \in \mathcal{D}_{t,n}$. We introduce a pseudo-OOS (out-of-source) dataset, $\mathcal{D}_s^{(od)}$ further in this section.

3.4.1 Goal task source pre-training (Fig. 4A). We train the backbone h and goal classifier f_g with source data $\mathcal{D}_s$ and stickered-source data $\mathcal{D}_{s,n}$:

$$\min_{\theta_h, \theta_{f_g}} \mathop{\mathbb{E}}_{(x,y)\in\mathcal{D}_s\cup\mathcal{D}_{s,n}} [\mathcal{L}_{s,g}]; \;\; \mathcal{L}_{s,g} = \mathcal{L}_{ce}(f_g \circ h(x), y) \tag{7}$$

Here, θ_h and θ_{f_g} are the parameters of h and f_g, $\mathcal{L}_{ce}$ is the cross-entropy loss, y is the goal task label, and expectation is implemented by sampling mini-batches.

3.4.2 Sticker task source pre-training (Fig. 4A). We pretrain the sticker classifier f_n while inculcating the ability to reject samples out of the source distribution. Specifically, f_n predicts a $(|\mathcal{C}_n| + 1)$-sized vector and is trained to classify *out-of-source* (OOS) samples to the $(|\mathcal{C}_n| + 1)^{\text{th}}$ class.

Insight 6. *The OOS node in the sticker classifier implicitly behaves as a domain discriminator from adversarial alignment methods. Minimizing the OOS probability only for the target data aligns the target with the source.*

Remarks. In source training, the OOS objective forces the sticker classifier to discriminate between source and OOS samples. This is done with the intuition that OOS samples simulate the role of target samples in adversarial alignment methods. This domain discriminatory knowledge will support future source-free target alignment. Concretely, the shared backbone can be adapted to the target, by minimizing OOS probability for target samples, as source knowledge is preserved by freezing f_g. Thus, we require OOS data to prepare f_n for adaptation.

Obtaining the OOS dataset. A naive approach is to use a dataset unrelated to the goal task label set. Conversely, we devise a pseudo-OOS dataset using only already available source samples. Mitsuzumi *et al.* [32] show that, beyond a

Table 1. Single-Source Domain Adaptation (SSDA) on Office-Home benchmarks. SF indicates *source-free* adaptation.

Method	SF	Office-Home												
		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
FixBi [33]	✗	58.1	77.3	80.4	67.7	79.5	78.1	65.8	57.9	81.7	76.4	62.9	86.7	72.7
SENTRY[41]	✗	61.8	77.4	80.1	66.3	71.6	74.7	66.8	63.0	80.9	74.0	66.3	84.1	72.2
SCDA [24]	✗	60.7	76.4	82.8	69.8	77.5	78.4	68.9	59.0	82.7	74.9	61.8	84.5	73.1
SHOT [26]	✓	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
A^{2}Net [58]	✓	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
GSFDA [60]	✓	57.9	78.6	81.0	66.7	77.2	77.2	65.6	56.0	82.2	72.0	57.8	83.4	71.3
CPGA [42]	✓	59.3	78.1	79.8	65.4	75.5	76.4	65.7	**58.0**	81.0	72.0	64.4	83.3	71.6
NRC [59]	✓	57.7	80.3	82.0	68.1	79.8	78.6	65.3	56.4	83.0	71.0	58.6	85.6	72.2
SHOT++[27]	✓	57.9	79.7	82.5	68.5	79.6	79.3	68.5	57.0	**83.0**	73.7	60.7	84.9	73.0
Ours	✓	**61.0**	**80.4**	**82.5**	**69.1**	**79.9**	**79.5**	**69.1**	57.8	82.7	**74.5**	**65.1**	**86.4**	**74.0**

certain grid size, shuffling grid patches makes the domain unrecognizable. Hence, we generate a pseudo-OOS dataset by shuffling grid patches of source images.

We also add stickers to shuffled images, at random, to further instill differences between source and pseudo-OOS (see Suppl). Formally, $(x_s^{(od)}, y_s^{(od)}) \in \mathcal{D}_s^{(od)}$ where $y_s^{(od)}$ denotes OOS category *i.e.* $(|\mathcal{C}_n|+1)^{\text{th}}$ category of f_n.

We train only the sticker classifier f_n, keeping backbone h and goal classifier f_g frozen, using cross-entropy loss $\mathcal{L}_{ce}$. With $\mathcal{L}_{s,n} = \mathcal{L}_{ce}(f_n \circ h(x_{s,n}), y_n)$, the overall objective for stickered source data $\mathcal{D}_{s,n}$ and pseudo-OOS data $\mathcal{D}_s^{(od)}$ is,

$$\min_{\theta_{f_n}} \mathop{\mathbb{E}}_{\mathcal{D}_{s,n}}[\mathcal{L}_{s,n}] + \mathop{\mathbb{E}}_{\mathcal{D}_s^{(od)}}[\mathcal{L}_s^{(od)}]; \quad \text{where } \mathcal{L}_s^{(od)} = \mathcal{L}_{ce}(f_n \circ h(x_s^{(od)}), y_s^{(od)}) \tag{8}$$

3.4.3 Source-free target adaptation (Fig. 4B). For unsupervised goal task adaptation, we use the general self training loss $\mathcal{L}_{st}$ and diversity loss $\mathcal{L}_{div}$ [26]. See Suppl. for more details. The goal task objective is given in Eq. 9 (left),

$$\min_{\theta_h} \mathop{\mathbb{E}}_{\mathcal{D}_t \cup \mathcal{D}_{t,n}}[\mathcal{L}_{st}+\mathcal{L}_{div}]; \;\; \text{and} \;\; \min_{(\theta_h, \theta_{f_n})} \mathop{\mathbb{E}}_{\mathcal{D}_{t,n}}[\mathcal{L}_{t,n}]; \;\; \mathcal{L}_{t,n} = \mathcal{L}_{ce}(f_n \circ h(x_{t,n}), y_n) \tag{9}$$

The goal classifier f_g is frozen to preserve its inductive bias and only the backbone h is updated for both original and stickered samples in Eq. 9 (left).

For subsidiary supervised sticker adaptation, we use a simple cross-entropy loss with sticker labels. We implicitly minimize OOS probability by maximizing label class probability. We observe that this works well and explicit minimization of OOS probability is not required. As per Insight 6, *out-of-target* (OOT) samples are not required. Further, using OOT samples to update the backbone could be undesirable as discussed under Insight 2. The objective is given in Eq. 9 (right). Both backbone h and sticker classifier f_n are updated as the task is supervised.

Table 2. Multi-Source Domain Adaptation (MSDA) on DomainNet and Office-Home. We outperform *source-free* (denoted by SF) prior arts despite not using domain labels.

Method	SF	w/o Domain Labels	DomainNet							Office-Home				
			→C	→I	→P	→Q	→R	→S	Avg	→Ar	→Cl	→Pr	→Rw	Avg
WAMDA [1]	✗	✗	59.3	21.8	52.1	9.5	65.0	47.7	42.6	71.9	61.4	84.1	82.3	74.9
SImpAl$_{50}$ [53]	✗	✗	66.4	26.5	56.6	18.9	68.0	55.5	48.6	70.8	56.3	80.2	81.5	72.2
CMSDA [48]	✗	✗	70.9	26.5	57.5	21.3	68.1	59.4	50.4	71.5	67.7	84.1	82.9	76.6
DRT [25]	✗	✗	71.0	31.6	61.0	12.3	71.4	60.7	51.3	-	-	-	-	-
STEM [34]	✗	✗	72.0	28.2	61.5	25.7	72.6	60.2	53.4	-	-	-	-	-
Source-combine	✗	✓	57.0	23.4	54.1	14.6	67.2	50.3	44.4	58.0	57.3	74.2	77.9	66.9
SHOT [26]-Ens	✓	✗	58.6	25.2	55.3	15.3	**70.5**	52.4	46.2	72.2	59.3	82.8	82.9	74.3
DECISION [2]	✓	✗	61.5	21.6	54.6	**18.9**	67.5	51.0	45.9	74.5	59.4	84.4	83.6	75.5
SHOT++ [27]	✓	✗	-	-	-	-	-	-	-	73.1	61.3	84.3	84.0	75.7
CAiDA [8]	✓	✗	-	-	-	-	-	-	-	**75.2**	60.5	84.7	84.2	76.2
NRC [59]	✓	✓	65.8	24.1	56.0	16.0	69.2	53.4	47.4	70.6	60.0	84.6	83.5	74.7
Ours	✓	✓	**70.3**	**25.7**	**57.3**	17.1	69.9	**57.1**	**49.6**	75.1	**64.1**	**86.6**	**84.4**	**77.6**

4 Experiments

We provide the implementation details of our experiments and thoroughly evaluate our approach w.r.t. state-of-the-art prior works across multiple settings. Unless mentioned, *Ours* implies sticker classification as the subsidiary task.

4.1 Experimental setup

Datasets. We evaluate on four standard DA benchmarks; Office-31 [44], Office-Home [54], VisDA [40], and DomainNet [39]. See Suppl. for more details.
Implementation details. We use a ResNet-101 [14] backbone for VisDA, and ResNet-50 for other benchmarks. We employ the same network design as SHOT [26]. For the subsidiary classifier, we use the same architecture after ResLayer-3. The number of sticker classes is 10. See Suppl. for more details.

4.2 Discussion

a) Single Source Domain Adaptation (SSDA). We compare with prior source-free SSDA works in Table 1 on Office-Home. We achieve *state-of-the-art* results exceeding the source-free SHOT++ and non-source-free SCDA [24] by 1% and 0.9% respectively. See Suppl. for Office-31 and VisDA results.
b) Multi Source Domain Adaptation (MSDA). In Table 2, we compare with the source-only baseline (*source-combine*) and source-free works. Even without domain labels, our approach achieves *state-of-the-art* results, even w.r.t. non-source-free works on Office-Home (+1%). On DomainNet, we outperform source-free works (+2.2%) with comparable results to non-source-free works.
c) Evaluating the subsidiary DA suitability criteria. We empirically evaluate DSM and TSM for our sticker-based tasks as well as existing tasks from self-supervised literature in Fig. 5A, 5B. Compared to patch location [51] and image rotation [31], sticker location and sticker rotation tasks exhibit higher

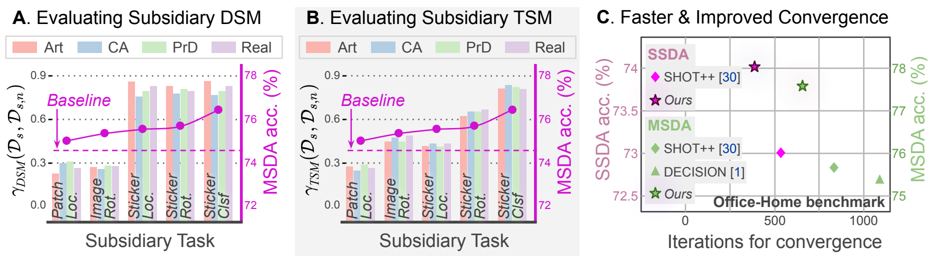

Fig. 5. We observe higher **A.** domain similarity (DSM) and **B.** task similarity (TSM) for our sticker-based tasks compared to existing subsidiary tasks like patch-location and image-rotation. This correlates with the better MSDA performance of sticker-based tasks on Office-Home and validates our criteria (Definition 1). **C.** Faster and improved convergence w.r.t. prior source-free works on both SSDA and MSDA for Office-Home.

DSM and thus, are more suitable with better adaptation performance (Table 3). However, sticker classification task is the most suitable due to its higher TSM as shape is the primary discriminative features, same as in goal task. We observe a positive correlation between DA performance and both DSM and TSM, which empirically verifies our suitability criteria. In Table 3, we also compare dense output based tasks like colorization and inpainting, which give marginal gains.

d) Faster and improved convergence. Fig. 5C illustrates our better and faster convergence w.r.t. source-free prior arts for both SSDA and MSDA. The hypothesis space for concurrent subsidiary supervised DA and unsupervised goal task DA, $\mathcal{H}_{g,n}$, is a subset of the hypothesis space for only unsupervised goal task DA, $\mathcal{H}_g^{(uns)}$. Thus, we achieve faster convergence. Further, as per Insight 1, lower domain discrepancy leads to lower target error *i.e.* improved convergence.

e) Compatibility with non-source-free DA. In Table 4, we evaluate the compatibility of concurrent subsidiary supervised DA with existing non-source-free SSDA techniques [10,30,52]. MSDA results are obtained by combining the multiple sources for each target. Compared to the original reported results, all four perform better with our proposed subsidiary DA. Note that our non-source-free variant outperforms these results (#7 in Table 5).

Table 3. Subsidiary task comparisons on **Office-Home** for source-free DA. Here, baseline is same as #3 in Table 5.

Method	SSDA	MSDA
Baseline (B)	66.2	74.3
B + inpainting	66.3	74.5
B + colorization	66.8	74.7
B + jigsaw	67.0	74.8
B + patch-loc	67.6	75.0
B + rotation	67.9	75.4
B + sticker-loc	68.8	75.5
B + sticker-rot	69.0	75.7
B + sticker-clsf	**69.7**	**76.2**

4.2.1 Ablation Study. Below, we discuss a thorough ablation study.

a) Effect of subsidiary supervised DA and OOS node. In Table 5, we compare the baseline *i.e.* only unsupervised goal task DA (#3) with the addition of only OOS classifier (#4). Here, a binary classifier is used for OOS detection. We

Table 4. Evaluating compatibility of subsidiary DA with non-source-free DA works on Office-Home. SSDA and MSDA indicate single-source and multi-source DA.

Method	Office-Home	
	SSDA	MSDA
CDAN [30]	65.8	69.4
+ *Subsidiary-DA*	**67.1**	**71.2**
SRDC [52]	71.3	73.1
+ *Subsidiary-DA*	**71.9**	**75.2**
FixBi [33]	72.7	-
+ *Subsidiary-DA*	**73.7**	-
CMSDA [48]	-	76.6
+ *Subsidiary-DA*	-	**78.1**

Table 5. Ablation analysis. Here, *sticker-w-OOS-clsf* denotes learning with all the proposed components unlike in *only-OOS-clsf* (all losses except $\mathcal{L}_{s,n}, \mathcal{L}_{t,n}$) and *only-sticker-clsf* (all losses except $\mathcal{L}_s^{(od)}$). SF denotes source-free constraint.

#	Variation	SF	Office-Home	
			SSDA	MSDA
1.	Source-only baseline	-	60.2	66.9
2.	+ *sticker-w-OOS-clsf*	-	61.9	71.4
3.	Adaptation baseline (B)	✓	66.2	74.3
4.	B + *only-OOS-clsf*	✓	67.0	74.9
5.	B + *only-sticker-clsf*	✓	69.7	76.2
6.	B + *sticker-w-OOS-clsf*	✓	73.1	77.6
7.	B + *sticker-w-OOS-clsf*	✗	**74.5**	**78.3**

observe gains of 0.8% and 0.6% for SSDA and MSDA respectively. This indicates that only OOS helps, but subsidiary classifier is essential for further improvements. Next, we compare the baseline (#3) with concurrent goal-subsidiary DA without using OOS (#5). We observe an improvement of 3.5% and 1.9% for SSDA and MSDA. Adding the OOS objective to the subsidiary supervised DA (#6 vs. #4) improves the source-target alignment as explained in Insight 6, resulting in improvements of 3.1% and 1.4% for SSDA and MSDA.

b) Subsidiary-goal task similarity. As per Insight 3, higher goal-subsidiary task similarity is important for effective learning of both tasks. Thus, in Table 5, we compare the source-only baseline (#1) with only subsidiary supervised DA without goal task target adaptation (#2). We observe gains of 1.7% and 1.3% for SSDA and MSDA respectively. This illustrates the positive correlation between sticker classification and goal task even when target goal losses are not used.

5 Conclusion

In this work, we introduced concurrent subsidiary supervised DA for a pretext-like task to aid the unsupervised goal task DA. We provide theoretical insights to analyze the effect of subsidiary supervised DA on the domain discrepancy and consequently on the goal task adaptation. Based on the insights, we introduce a subsidiary DA suitability criteria to determine DA-assistive subsidiary tasks that improve the goal task DA performance. We also propose a novel sticker intervention based pretext task that follows our criteria. The proposed approach outperforms prior state-of-the-art source-free SSDA and MSDA works on four standard benchmarks, establishing the usefulness of our approach.

Acknowledgments. This work was supported by MeitY (Ministry of Electronics and Information Technology) project (No. 4(16)2019-ITEA), Govt. of India and a research grant by Google.

References

1. Aggarwal, S., Kundu, J.N., Babu, R.V., Chakraborty, A.: WAMDA: Weighted alignment of sources for multi-source domain adaptation. In: BMVC (2020) 12
2. Ahmed, S.M., Raychaudhuri, D.S., Paul, S., Oymak, S., Roy-Chowdhury, A.K.: Unsupervised multi-source domain adaptation without access to source data. In: CVPR (2021) 12
3. Ben-David, S., Blitzer, J., Crammer, K., Pereira, F.: Analysis of representations for domain adaptation. In: NeurIPS (2006) 1, 2, 5
4. Bousmalis, K., Silberman, N., Dohan, D., Erhan, D., Krishnan, D.: Unsupervised pixel-level domain adaptation with generative adversarial networks. In: CVPR (2017) 4
5. Carlucci, F.M., D'Innocente, A., Bucci, S., Caputo, B., Tommasi, T.: Domain generalization by solving jigsaw puzzles. In: CVPR (2019) 1, 3, 4, 7, 8
6. Chen, Y.H., Chen, W.Y., Chen, Y.T., Tsai, B.C., Frank Wang, Y.C., Sun, M.: No more discrimination: Cross city adaptation of road scene segmenters. In: ICCV (2017) 1
7. Doersch, C., Gupta, A., Efros, A.A.: Unsupervised visual representation learning by context prediction. In: ICCV (2015) 4
8. Dong, J., Fang, Z., Liu, A., Sun, G., Liu, T.: Confident anchor-induced multi-source free domain adaptation. In: NeurIPS (2021) 12
9. Feng, Z., Xu, C., Tao, D.: Self-supervised representation learning from multi-domain data. In: ICCV (2019) 4
10. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., Lempitsky, V.: Domain-adversarial training of neural networks. The Journal of Machine Learning Research **17**(1), 2096–2030 (2016) 1, 5, 9, 13
11. Ghifary, M., Kleijn, W.B., Zhang, M., Balduzzi, D.: Domain generalization for object recognition with multi-task autoencoders. In: ICCV (2015) 4
12. Ghifary, M., Kleijn, W.B., Zhang, M., Balduzzi, D., Li, W.: Deep reconstruction-classification networks for unsupervised domain adaptation. In: ECCV (2016) 4
13. Gidaris, S., Singh, P., Komodakis, N.: Unsupervised representation learning by predicting image rotations. In: ICLR (2018) 1, 2, 4
14. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016) 12
15. Jiaolong, X., Liang, X., López, A.M.: Self-supervised domain adaptation for computer vision tasks. IEEE Access **7**, 156694–156706 (2019) 4, 8
16. Kim, D., Saito, K., Oh, T.H., Plummer, B.A., Sclaroff, S., Saenko, K.: CDS: Cross-domain self-supervised pre-training. In: ICCV (2021) 4
17. Kundu, J.N., Kulkarni, A., Bhambri, S., Mehta, D., Kulkarni, S., Jampani, V., Babu, R.V.: Balancing discriminability and transferability for source-free domain adaptation. In: ICML (2022) 3
18. Kundu, J.N., Kulkarni, A., Singh, A., Jampani, V., Babu, R.V.: Generalize then adapt: Source-free domain adaptive semantic segmentation. In: ICCV (2021) 4
19. Kundu, J.N., Venkat, N., M V, R., Babu, R.V.: Universal source-free domain adaptation. In: CVPR (2020) 4, 9
20. Kundu, J.N., Venkat, N., Revanur, A., V, R.M., Babu, R.V.: Towards inheritable models for open-set domain adaptation. In: CVPR (2020) 4
21. Kundu, J.N., Venkatesh, R.M., Venkat, N., Revanur, A., Babu, R.V.: Class-incremental domain adaptation. In: ECCV (2020) 3

22. Larsson, G., Maire, M., Shakhnarovich, G.: Colorization as a proxy task for visual understanding. In: CVPR (2017) 3, 4, 8
23. Li, R., Jiao, Q., Cao, W., Wong, H.S., Wu, S.: Model adaptation: Unsupervised domain adaptation without source data. In: CVPR (2020) 3, 4, 9
24. Li, S., Xie, M., Lv, F., Liu, C.H., Liang, J., Qin, C., Li, W.: Semantic concentration for domain adaptation. In: ICCV (2021) 11, 12
25. Li, Y., Yuan, L., Chen, Y., Wang, P., Vasconcelos, N.: Dynamic transfer for multi-source domain adaptation. In: CVPR (2021) 12
26. Liang, J., Hu, D., Feng, J.: Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In: ICML (2020) 3, 4, 11, 12
27. Liang, J., Hu, D., Wang, Y., He, R., Feng, J.: Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer. IEEE Transactions on Pattern Analysis and Machine Intelligence (2021) 4, 11, 12
28. Liu, Y., Zhang, W., Wang, J.: Source-free domain adaptation for semantic segmentation. In: CVPR (2021) 4
29. Long, M., Cao, Y., Wang, J., Jordan, M.: Learning transferable features with deep adaptation networks. In: ICML (2015) 5
30. Long, M., Cao, Z., Wang, J., Jordan, M.I.: Conditional adversarial domain adaptation. In: NeurIPS (2017) 1, 13, 14
31. Mishra, S., Saenko, K., Saligrama, V.: Surprisingly simple semi-supervised domain adaptation with pretraining and consistency. In: BMVC (2021) 1, 2, 3, 8, 9, 12
32. Mitsuzumi, Y., Irie, G., Ikami, D., Shibata, T.: Generalized domain adaptation. In: CVPR (2021) 7, 10
33. Na, J., Jung, H., Chang, H.J., Hwang, W.: FixBi: Bridging domain spaces for unsupervised domain adaptation. In: CVPR (2021) 11, 14
34. Nguyen, V.A., Nguyen, T., Le, T., Tran, Q.H., Phung, D.: STEM: An approach to multi-source domain adaptation with guarantees. In: ICCV (2021) 12
35. Noroozi, M., Favaro, P.: Unsupervised learning of visual representations by solving jigsaw puzzles. In: ECCV (2016) 1, 2, 4
36. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018) 4
37. Parkhi, O.M., Vedaldi, A., Zisserman, A., Jawahar, C.V.: Cats and dogs. In: CVPR (2012) 2
38. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A.: Context encoders: Feature learning by inpainting. In: CVPR (2016) 3, 4, 8
39. Peng, X., Bai, Q., Xia, X., Huang, Z., Saenko, K., Wang, B.: Moment matching for multi-source domain adaptation. In: ICCV (2019) 12
40. Peng, X., Usman, B., Kaushik, N., Hoffman, J., Wang, D., Saenko, K.: VisDA: The visual domain adaptation challenge. arXiv preprint arXiv:1710.06924 (2017) 12
41. Prabhu, V., Khare, S., Kartik, D., Hoffman, J.: SENTRY: Selective entropy optimization via committee consistency for unsupervised domain adaptation. In: ICCV (2021) 11
42. Qiu, Z., Zhang, Y., Lin, H., Niu, S., Liu, Y., Du, Q., Tan, M.: Source-free domain adaptation via avatar prototype generation and adaptation. In: IJCAI (2021) 11
43. Ren, Z., Lee, Y.J.: Cross-domain self-supervised multi-task feature learning using synthetic imagery. In: CVPR (2018) 4
44. Saenko, K., Kulis, B., Fritz, M., Darrell, T.: Adapting visual category models to new domains. In: ECCV (2010) 12
45. Saito, K., Kim, D., Sclaroff, S., Saenko, K.: Universal domain adaptation through self supervision. In: NeurIPS (2020) 4

46. Saito, K., Yamamoto, S., Ushiku, Y., Harada, T.: Open set domain adaptation by backpropagation. In: ECCV (2018) 4
47. Salman, H., Ilyas, A., Engstrom, L., Kapoor, A., Madry, A.: Do adversarially robust imagenet models transfer better? In: NeurIPS (2020) 5
48. Scalbert, M., Vakalopoulou, M., Couzini'e-Devy, F.: Multi-source domain adaptation via supervised contrastive learning and confident consistency regularization. In: BMVC (2021) 12, 14
49. Sivaprasad, P.T., Fleuret, F.: Uncertainty reduction for model adaptation in semantic segmentation. In: CVPR (2021) 4
50. Sun, B., Saenko, K.: Deep CORAL: Correlation alignment for deep domain adaptation. In: ECCV (2016) 5, 9
51. Sun, Y., Tzeng, E., Darrell, T., Efros, A.A.: Unsupervised domain adaptation through self-supervision. arXiv preprint arXiv:1909.11825 (2019) 1, 3, 4, 8, 9, 12
52. Tang, H., Chen, K., Jia, K.: Unsupervised domain adaptation via structurally regularized deep clustering. In: CVPR (2020) 13, 14
53. Venkat, N., Kundu, J.N., Singh, D.K., Revanur, A., Babu, R.V.: Your classifier can secretly suffice multi-source domain adaptation. In: NeurIPS (2020) 12
54. Venkateswara, H., Eusebio, J., Chakraborty, S., Panchanathan, S.: Deep hashing network for unsupervised domain adaptation. In: CVPR (2017) 12
55. Wallace, B., Hariharan, B.: Extending and analyzing self-supervised learning across domains. In: ECCV (2020) 2
56. Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., Summers, R.: ChestX-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In: CVPR (2017) 2
57. Wu, Z., Xiong, Y., Yu, S.X., Lin, D.: Unsupervised feature learning via nonparametric instance discrimination. In: CVPR (2018) 4
58. Xia, H., Zhao, H., Ding, Z.: Adaptive adversarial network for source-free domain adaptation. In: ICCV (2021) 11
59. Yang, S., Wang, Y., van de Weijer, J., Herranz, L., Jui, S.: Exploiting the intrinsic neighborhood structure for source-free domain adaptation. In: NeurIPS (2021) 4, 11, 12
60. Yang, S., Wang, Y., van de Weijer, J., Herranz, L., Jui, S.: Generalized source-free domain adaptation. In: ICCV (2021) 11
61. You, K., Long, M., Cao, Z., Wang, J., Jordan, M.I.: Universal domain adaptation. In: CVPR (2019) 4
62. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: ECCV (2016) 4, 8
63. Zhang, R., Isola, P., Efros, A.A.: Split-brain autoencoders: Unsupervised learning by cross-channel prediction. In: CVPR (2017) 4
64. Zhao, H., Combes, R.T.D., Zhang, K., Gordon, G.: On learning invariant representations for domain adaptation. In: ICML (2019) 2, 5