UNIVERSITYOF
BIRMINGHAM

Contrastive vicinal space for unsupervised domain adaptation

Na, Jaemin; Han, Dongyoon; Chang, Hyung Jin; Hwang, Wonjun

Link to publication on Research at Birmingham portal

Contrastive Vicinal Space for Unsupervised Domain Adaptation

Jaemin Na[1] iD, Dongyoon Han[2] iD, Hyung Jin Chang[3] iD, and Wonjun Hwang[1] iD

[1]Ajou University, Korea   [2]NAVER AI Lab   [3]University of Birmingham, UK
osial46@ajou.ac.kr, dongyoon.han@navercorp.com,
h.j.chang@bham.ac.uk, wjhwang@ajou.ac.kr

Abstract. Recent unsupervised domain adaptation methods have utilized vicinal space between the source and target domains. However, the equilibrium collapse of labels, a problem where the source labels are dominant over the target labels in the predictions of vicinal instances, has never been addressed. In this paper, we propose an instance-wise minimax strategy that minimizes the entropy of high uncertainty instances in the vicinal space to tackle the stated problem. We divide the vicinal space into two subspaces through the solution of the minimax problem: contrastive space and consensus space. In the contrastive space, inter-domain discrepancy is mitigated by constraining instances to have contrastive views and labels, and the consensus space reduces the confusion between intra-domain categories. The effectiveness of our method is demonstrated on public benchmarks, including Office-31, Office-Home, and VisDA-C, achieving state-of-the-art performances. We further show that our method outperforms the current state-of-the-art methods on PACS, which indicates that our instance-wise approach works well for multi-source domain adaptation as well. Code is available at https://github.com/NaJaeMin92/CoVi.

Keywords: Unsupervised Domain Adaptation, Equilibrium Collapse, Contrastive Vicinal Space

1 Introduction

Unsupervised domain adaptation (UDA) aims to adapt a model trained on a labeled source domain to an unlabeled target domain. One of the most important problems to solve in UDA is the domain shift [42] (*i.e.,* distribution shift) problem. The domain shift arises from the change in the data distribution between the training domain (*i.e.,* source domain) of an algorithm and the test domain encountered in a practical application (*i.e.,* target domain). Although recent UDA studies [33,54,49] have shown encouraging results, a large domain shift is still a significant obstacle.

One recent paradigm to address the large domain shift problem is to leverage intermediate domains between the source and target domains instead of direct
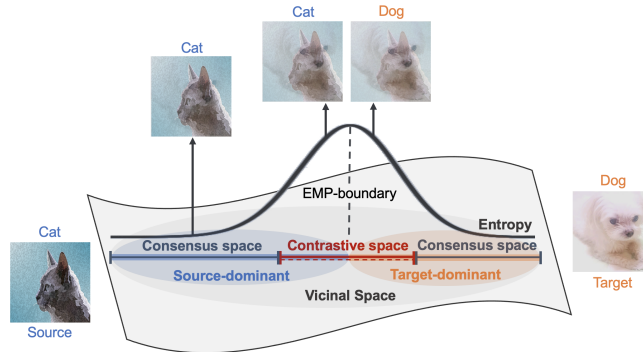
Fig. 1: **Overview.** Vicinal space between the source and target domains is divided into contrastive space and consensus space. Our methodology alleviates inter-domain discrepancy in the contrastive space and simultaneously resolves intra-domain categorical confusion in the consensus space.

domain adaptation. Recent studies [15,5] inspired by generative adversarial networks [16] (GANs) generate instances of intermediate domains to bridge the source and target domains. Moreover, [10,3] learn domain-invariant representations by borrowing only the concept of adversarial training. Meanwhile, with the development of data augmentation techniques, many approaches have emerged built on data augmentation to construct the intermediate spaces. Recent studies [33,51,48] have shown promising results by grafting Mixup augmentation [55] to the domain adaptation task. These studies use inter-domain mixup to efficiently overcome the domain shift problem by utilizing vicinal instances between the source and target domains. However, none of them consider leveraging the predictions of the vicinal instances in the perspective of self-training [24].

Self-training is the straightforward approach that uses self-predictions of a model to train itself. Semi-supervised learning methods [24,36] leverage a model's predictions on unlabeled data to obtain additional information used during training as their supervision. In particular, unsupervised domain adaptation methods [17,43,38] have shown that pseudo-label for the target domain can play an important role in alleviating the domain shift problem.

In this work, we introduce a new **Co**ntrastive **Vi**cinal space-based (CoVi) algorithm that leverages vicinal instances from the perspective of self-training [24]. In vicinal space, we observe that the source label is generally dominant over the target label before applying domain adaptation. In other words, even if vicinal instances consist of a higher proportion of target instances than source instances (*i.e.,* target-dominant instances), their one-hot predictions are more likely to be source labels (*i.e.,* source-dominant labels). We define this phenomenon as an **equilibrium collapse of labels** between vicinal instances. We also discover that the entropy of the predictions is maximum at the points where the equilibrium collapse of labels occurs. Hence, we aim to find and address the points where the entropy is maximized between the vicinal instances. Inspired by the minimax strategy [12], we present *EMP-Mixup*, which minimizes the entropy for the **entropy maximization point (EMP)**. Our *EMP-Mixup* adaptively

adjusts the Mixup ratio according to the combinations of source and target instances through training.

As depicted in Figure 1, we further leverage the EMP as a boundary (*i.e.*, EMP-boundary) to divide the vicinal space into source-dominant and target-dominant spaces. Here, the vicinal instances of the source-dominant space have source labels as their predicted top-1 label. Similarly, the vicinal instances of target-dominant space have target labels as their top-1 label. Taking advantage of these properties, we configure two specialized subspaces to reduce inter-domain and intra-domain discrepancy simultaneously.

First, we construct a **contrastive space** around the EMP-boundary to ensure that the vicinal instances have contrastive views: source-dominant and target-dominant views. Since the contrastive views share the same combination of source and target instances, they should have the same top-2 labels containing the source and target labels. In addition, under our constraints, the two contrastive views have opposite order of the first and second labels in the top-2 labels. Inspired by consistency training [44,1], we propose to impose consistency on predictions of the two contrastive views. Specifically, we mitigate inter-domain discrepancy by solving a *"swapped" prediction problem* where we predict the top-2 labels of a contrastive view from the other contrastive view.

Second, we constrain a **consensus space** outside of the contrastive space to alleviate the categorical confusion within the intra-domain. In this space, we generate target-dominant vicinal instances utilizing multiple source instances as a perturbation to a single target instance. Here, the role of the source instances is not to learn classification information of the source domain but to confuse the predictions of the target instances. We can ensure consistent and robust predictions for target instances by enforcing label consensus among the multiple target-dominant vicinal instances to a single target label.

We perform extensive ablation studies for a detailed analysis of the proposed methods. In particular, we achieve comparable performance to the recent state-of-the-art methods in standard unsupervised domain adaptation benchmarks such as Office-31 [37], Office-Home [46], and VisDA-C [35]. Furthermore, we validate the superiority of our instance-wise approach on the PACS [26] dataset for multi-source domain adaptation. Overall, we make the following contributions:

- This is the first study in UDA to leverage the vicinal space from the perspective of self-training. We shed light on the problem of the equilibrium collapse of labels in the vicinal space and propose a minimax strategy to handle it.
- We alleviate inter-domain and intra-domain confusions simultaneously by dividing the vicinal space into contrastive and consensus spaces.
- Our method achieves state-of-the-art performance and is further validated through extensive ablation studies.

2   Related Work

Unsupervised domain adaptation. One of the representative domain adaptation approaches [13,50] is learning a domain-invariant representation by aligning

the global distribution between the source and target domains. Of particular interest, Xie *et al.* [50] presented a moving semantic transfer network that aligns labeled source centroids and pseudo-labeled target centroids to learn semantic representations for unlabeled target data. Following [33,6,17], we adopt this simple but efficient method as our baseline.

Our work is also related to the domain adaptation approaches that consider the inter-domain and intra-domain gap together. Kang *et al.* [23] proposed to minimize the intra-class discrepancy and maximize the inter-class discrepancy to perform class-aware domain alignment. Pan *et al.* [34] presented a semantic segmentation method that minimizes both inter-domain and intra-domain gaps. Unlike these methods, we introduce a practical approach that uses two specialized spaces to reduce inter-domain and intra-domain discrepancy for each.

Mixup augmentation. Mixup [55] is a data-agnostic and straightforward augmentation using a linear interpolation between two data instances. The Mixup has been applied to various tasks and shown to improve the robustness of neural networks. The recent semi-supervised learning methods [2,44,1] efficiently utilized Mixup to leverage unlabeled data. Meanwhile, several domain adaptation methods [48,51,33] with Mixup were proposed to alleviate the domain-shift problem successfully. Xu *et al.* [51] and Wu *et al.* [48] showed promising results using inter-domain Mixup between source and target domains. Recently, Na *et al.* [33] achieved a significant performance gain by using two networks trained with two fixed Mixup ratios. Moreover, the latest studies [18,56,31] suggested adaptive Mixup techniques instead of using manually designed interpolation policies. For example, Zhu *et al.* [56] introduced a more advanced interpolation technique that seeks the Wasserstein barycenter between two instances and proposed an adaptive Mixup. Mai *et al.* [31] introduced a meta-learning-based optimization strategy for dynamically learning the interpolation policy in semi-supervised learning. However, unsupervised domain adaptation methods still count on hand-tuned or random interpolation policies. In this work, we derive the Mixup ratio according to the convex combinations of source and target instances.

Consistency training. Consistency training is one of the promising components for leveraging unlabeled data, which enforces a model to produce similar predictions of original and perturbed instances. The recent semi-supervised learning methods [44,1,2] utilize unlabeled data by assuming that the model should output similar predictions when fed perturbed versions of the same instance. Berthelot *et al.* [2] applied augmentations several times for each unlabeled instance and averaged them to produce guessed labels. In ReMixMatch [1], Berthelot *et al.* used the model's prediction for a weakly-augmented instance as the guessed label for multiple strongly-augmented variants of the same instance. Recently, Sohn *et al.* [44] encouraged predictions from strongly-augmented instances to match pseudo-labels generated from weakly-augmented instances. Although effective, these methods rely on augmentation techniques such as random augmentation, AutoAugment [8], RandAugment [9], and CTAugment [1]. By contrast, our method is free from these augmentation techniques and does not require carefully selected combinations of augmentations. We solely leverage
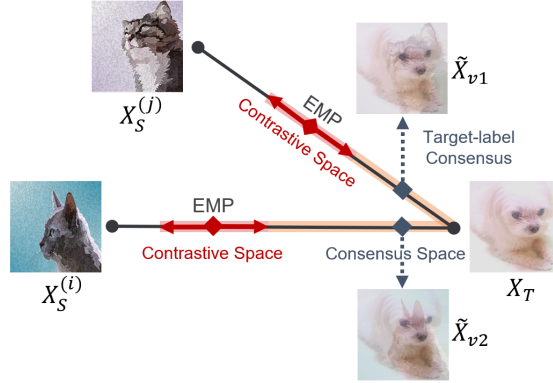
Fig. 2: **Schematic illustration of CoVi.** The *EMP-Mixup* finds the most confusing point (*i.e.*, EMP) among vicinal instances. CoVi then learns through top-k contrastive predictions from contrastive views in the contrastive space determined by the EMP. In the consensus space, we achieve a target-label consensus with perturbations of the source instances.

mixup augmentation [55] to generate the vicinal spaces and achieve the effect of consistency training.

3   Methodology

CoVi introduces three techniques to leverage the vicinal space between the source and target domains: i) EMP-Mixup, ii) contrastive views and labels, and iii) a label-consensus. An overall depiction of CoVi is in Figure 2.

3.1   Preliminaries

Notation. We denote a mini-batch of m-images as $\mathcal{X}$, corresponding labels as $\mathcal{Y}$, and extracted features from $\mathcal{X}$ as $\mathcal{Z}$. Specifically, $\mathcal{X}_\mathcal{S} \subset \mathbb{R}^{m \times i}$ and $\mathcal{Y}_\mathcal{S} \subset \{0,1\}^{m \times n}$ denote the mini-batches of source instances and their corresponding one-hot labels, respectively. Here, n denotes the number of classes and $i = c \cdot h \cdot w$, where c denotes the channel size, and h and w denote the height and width of the image instances, respectively. Similarly, the mini-batch of unlabeled target instances is $\mathcal{X}_\mathcal{T} \subset \mathbb{R}^{m \times i}$. Our model consists of the following subcomponents: an *encoder* f_θ, a *classifier* h_θ, and an *EMP-learner* g_ϕ.

Mixup. The *Mixup* augmentation [55] based on the Vicinal Risk Minimization (VRM) [7] principle exploits virtual instances constructed with the linear interpolation of two instances. These vicinal instances can benefit unsupervised domain adaptation, which has no target domain labels. We define the inter-domain Mixup applied between the source and target domains as follows:

$$\tilde{\mathcal{X}}_\lambda = \lambda \cdot \mathcal{X}_\mathcal{S} + (1 - \lambda) \cdot \mathcal{X}_\mathcal{T}$$
$$\tilde{\mathcal{Y}}_\lambda = \lambda \cdot \mathcal{Y}_\mathcal{S} + (1 - \lambda) \cdot \hat{\mathcal{Y}}_\mathcal{T}, \tag{1}$$

where $\hat{\mathcal{Y}}_{\mathcal{T}}$ denotes the pseudo labels of the target instances and $\lambda \in [0,1]$ is the Mixup ratio. Then, the empirical risk for vicinal instances in the inter-domain Mixup is defined as follows:

$$\mathcal{R}_{\lambda} = \frac{1}{m} \sum_{i=1}^{m} \mathcal{H}[h(f(\tilde{\mathcal{X}}_{\lambda}^{(i)})), \tilde{\mathcal{Y}}_{\lambda}^{(i)}], \tag{2}$$

where $\mathcal{H}$ is a standard cross-entropy loss.

3.2   EMP-Mixup

In the vicinal space between the source and target domains, we make interesting observations on unsupervised domain adaptation.

Observation 1. *"The labels of the target domain are relatively recessive to the source domain labels."*

We investigate the dominance of the predicted top-1 labels between the source and target instances in vicinal instances. We find that the label dominance is balanced when the labels of both the source and target domains are provided (*i.e.,* supervised learning). In this case, the top-1 label of the vicinal instance is determined by the instance occupying a relatively larger proportion. However, in the UDA, where the label of the target domain is not given, the balance of label dominance is broken (*i.e.,* equilibrium collapse of labels). Indeed, we discover that source labels frequently represent vicinal instances even with a higher proportion of target instances than source instances.

Observation 2. *"Depending on the convex combinations of source and target instances, the label dominance is changed."*

Next, we observe that the label dominance is altered according to the convex combinations of instances. It implies that an instance-wise approach can be a key to solving the label equilibrium collapse problem. In addition, we discover that the entropy of the prediction is maximum at the point where the label dominance changes because the source and target instances become most confusing at this point (see Figures 4 and 5).

Based on these observations, we aim to capture and mitigate the most confusing points, which vary with the combination of instances. Inspired by [12,32,16], we introduce a minimax strategy to break through the worst-case risk [12] among the vicinal instances between the source and target domains. We minimize the worst risk by finding the **entropy maximization point (EMP)** among the vicinal instances. In order to estimate the EMPs, we introduce a small network, *EMP-learner*. This network aims to generate Mixup ratios that maximize the entropy of the encoder f_{θ} (*e.g., ResNet*) followed by a classifier h_{θ}.

Given $\mathcal{X}_{\mathcal{S}}$ and $\mathcal{X}_{\mathcal{T}}$, we obtain the instance features $\mathcal{Z}_{\mathcal{S}} = f_{\theta}(\mathcal{X}_{\mathcal{S}})$ and $\mathcal{Z}_{\mathcal{T}} = f_{\theta}(\mathcal{X}_{\mathcal{T}})$ from the encoder f_{θ}. Then, we pass the concatenated features $\mathcal{Z}_{\mathcal{S}} \oplus \mathcal{Z}_{\mathcal{T}}$ to the *EMP-learner* g_{ϕ}. Then, the *EMP-learner* produces the entropy maximization ratio λ^* that maximizes the entropy of the encoder f_{θ}. Formally, the Mixup ratios for our *EMP-Mixup* are defined as follows:

$$\lambda^* = \underset{\lambda \in [0,1]}{\arg\max} \, \mathcal{H}[h_{\theta}(f_{\theta}(\tilde{\mathcal{X}}_{\lambda}))], \tag{3}$$

where $\lambda = g_\phi(\mathcal{Z}_\mathcal{S} \oplus \mathcal{Z}_\mathcal{T})$ and $\mathcal{H}$ is the entropy loss.

Finally, we design the objective function for *EMP-learner* to **maximize the entropy** as follows:

$$\mathcal{R}_\lambda(\phi) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{H}[h(f(\tilde{\mathcal{X}}_\lambda^{(i)}))], \tag{4}$$

where $\mathcal{H}$ is the entropy loss. Note that we only update the parameter ϕ of the *EMP-learner*, not the parameter θ of the encoder and the classifier. With the worst-case ratio λ^*, *EMP-Mixup* **minimizes the worst-case risk** on vicinal instances as follows:

$$\mathcal{R}_{\lambda^*}(\theta) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{H}[h(f(\tilde{\mathcal{X}}_{\lambda^*}^{(i)})), \tilde{\mathcal{Y}}_{\lambda^*}^{(i)}], \tag{5}$$

where $\mathcal{H}$ is the standard cross-entropy loss.

It is noteworthy that our $\lambda^* = [\lambda_1, ..., \lambda_m]$ has different optimized ratios according to the combinations of the source and target instances within a mini-batch. Finally, *EMP-Mixup* minimizes the risk of vicinal instances from the viewpoint of the worst-case risk. The overall objective functions are defined as follows:

$$\mathcal{R}_{emp} = \mathcal{R}_{\lambda^*}(\theta) - \mathcal{R}_\lambda(\phi). \tag{6}$$

3.3 Contrastive Views and Labels

Observation 3. *"The dominant/recessive labels of the vicinal instances are switched at the EMP."*

Looking back to the previous observations, the label dominance depends on the convex combination of instances, and the point of change is the EMP. In other words, with the EMP as a boundary (*i.e., EMP-boundary*), the dominant/recessive label is switched between the source and target domains. It means that vicinal instances around the *EMP-boundary* should have source and target labels as their top-2 labels.

These observations and analyses lead us to design the concepts of contrastive views and contrastive labels. Owing to the *EMP-boundary*, we can divide the vicinal space into source-dominant and target-dominant space, as described in Figure 3. Specifically, we constrain the source-dominant and target-dominant spaces of the contrastive space to $\lambda^* - \omega < \lambda_{sd} < \lambda^*$ and $\lambda^* < \lambda_{td} < \lambda^* + \omega$, respectively. Here, ω is the margin of the ratio from the *EMP-boundary*, which is manually designed. Consequently, the source-dominant instances $\tilde{\mathcal{X}}_{sd}$ and target-dominant instances $\tilde{\mathcal{X}}_{td}$ have **contrastive views** of each other.

From the contrastive views, we focus on the top-2 labels for each prediction because we are only interested in the classes that correspond to the source and target instances, not the other classes. Here, we define a set of top-2 one-hot labels within a mini-batch as $\hat{\mathcal{Y}}_{[k=1]}$ and $\hat{\mathcal{Y}}_{[k=2]}$. Unlike a general Mixup that uses pure source and target labels (see Eq.1), we directly exploit the predicted
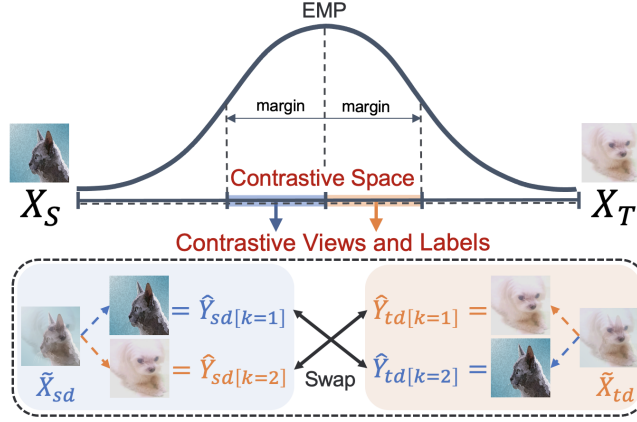
Fig. 3: **Contrastive views and labels.** (i) The **contrastive views** consist of a source-dominant view $\tilde{\mathcal{X}}_{sd}$ and a target-dominant view $\tilde{\mathcal{X}}_{td}$. (ii) The **contrastive labels** comprise the source-dominant label and target-recessive label from the top-2 predictions in the contrastive view $\tilde{\mathcal{X}}_{sd}$ (and vice versa).

labels from vicinal instances. In this case, for example, the labels for the instances of the target-dominant space are constructed as follows:

$$\hat{\mathcal{Y}}_{td} = \lambda_{td} \cdot \hat{\mathcal{Y}}_{td[k=1]} + (1 - \lambda_{td}) \cdot \hat{\mathcal{Y}}_{td[k=2]}. \tag{7}$$

Furthermore, we expand on this and propose a new concept of **contrastive labels**. We constrain the top-2 labels from the contrastive views as follows:

- $\hat{\mathcal{Y}}_{sd[k=1]}$ from $\tilde{\mathcal{X}}_{sd}$ and $\hat{\mathcal{Y}}_{td[k=2]}$ from $\tilde{\mathcal{X}}_{td}$ must be equal, as the predictions of the source instances.
- Similarly, $\hat{\mathcal{Y}}_{sd[k=2]}$ must be equal to $\hat{\mathcal{Y}}_{td[k=1]}$, as for the predictions of the target instances.

In other words, the dominant label $\hat{\mathcal{Y}}_{sd[k=1]}$ of $\tilde{\mathcal{X}}_{sd}$ and the recessive label $\hat{\mathcal{Y}}_{td[k=2]}$ of $\tilde{\mathcal{X}}_{td}$ must be the same as the source labels and vice versa. Note that our contrastive constraints are instance-level constraints that must be satisfied between any instances, regardless of the class category.

Consequently, we swap the top-2 contrastive labels between two contrastive views to learn from the predictions of the other view. By solving a "swapped" prediction problem, we enforce consistency to the top-2 contrastive labels obtained from contrastive views of the same source and target instance combinations.

According to the constraints, Eq.7 still holds when we swap the contrastive labels. Finally, the objective for our contrastive loss in target-dominant space is defined as follows:

$$\mathcal{R}_{td}(\theta) = \frac{1}{m} \sum_{i=1}^{m} \mathcal{H}[h(f(\tilde{\mathcal{X}}_{td}^{(i)})), \hat{\mathcal{Y}}_{td}^{(i)}], \tag{8}$$

where $\hat{\mathcal{Y}}_{td} = \lambda_{td} \cdot \hat{\mathcal{Y}}_{sd[k=2]} + (1 - \lambda_{td}) \cdot \hat{\mathcal{Y}}_{sd[k=1]}.$

Similarly, we define $\mathcal{R}_{sd}(\theta)$ in the source-dominant space and omit it for clarity. The overall objective functions for contrastive loss are defined as follows:

$$\mathcal{R}_{ct} = \mathcal{R}_{td}(\theta) + \mathcal{R}_{sd}(\theta). \tag{9}$$

3.4 Label Consensus

Even though the confusion between the source and target instances is crucial in the contrastive space, outside of the contrastive space (*i.e.*, **consensus space**), we pay more attention to the uncertainty of predictions within the intra-domain than inter-domain instances (see Figure 6). Here, we exploit multiple source instances to impose perturbations to target predictions rather than classification information for the source domain. It makes a model more robust to the target predictions by enforcing consistent predictions on the target instances even with the source perturbations.

We construct two randomly shuffled versions of the source instances within a mini-batch. We then apply Mixup with a single target mini-batch to obtain two different perturbed views v_1 and v_2. Here, we set the mixup ratio for the source instances sufficiently small since too strong perturbations can impair the target class semantics. We compute two softmax probabilities from the perturbed instances $\tilde{\mathcal{X}}_{v_1}$ and $\tilde{\mathcal{X}}_{v_2}$ using an encoder, followed by a classifier. Finally, we aggregate the softmax probabilities and yield a one-hot prediction $\hat{\mathcal{Y}}$.

We accomplish **target-label consensus** by assigning the label $\hat{\mathcal{Y}}$ to both versions of the perturbed target-dominant instances $\tilde{\mathcal{X}}_{v_1}$ and $\tilde{\mathcal{X}}_{v_2}$. Imposing consistency to differently perturbed instances for a single target label allows us to focus on categorical information for the target domain. The objective for label consensus on target instances can be defined as follows:

$$\mathcal{R}_{cs}(\theta) = \frac{1}{m} \sum_{i=1}^{m} [\mathcal{H}(h(f(\tilde{\mathcal{X}}_{v_1}^{(i)}), \hat{\mathcal{Y}}^{(i)})) + \mathcal{H}(h(f(\tilde{\mathcal{X}}_{v_2}^{(i)}), \hat{\mathcal{Y}}^{(i)}))], \tag{10}$$

where $\mathcal{H}$ is the cross-entropy loss.

Note that this approach is also applicable to source-dominant space, but we exclude it from the final loss as it does not significantly affect the performance.

4 Experiments

We evaluate our method on four popular benchmarks, including Office-31, Office-Home, VisDA-C, and PACS. Moreover, we validate our method in a multi-source domain adaptation scenario using the PACS dataset.

- Office-31 [37] contains 31 categories and 4,110 images in three domains: Amazon (A), Webcam (W), and DSLR (D). We verify our methodology in six domain adaptation tasks.

- Office-Home [46] consists of 64 categories and 15,500 images in four domains: Art (Ar), Clipart (Cl), Product (Pr), and Real-World (Rw).

Method	A→W	D→W	W→D	A→D	D→A	W→A	Avg.
MSTN* (Baseline) [50]	91.3	98.9	**100.0**	90.4	72.7	65.6	86.5
DWL (CVPR'21) [49]	89.2	99.2	**100.0**	91.2	73.1	69.8	87.1
DMRL (ECCV'20) [48]	90.8±0.3	99.0±0.2	**100.0±0.0**	93.4±0.5	73.0±0.3	71.2±0.3	87.9
ILA-DA (CVPR'21) [41]	95.72	99.25	**100.0**	93.37	72.10	75.40	89.3
MCC (ECCV'20) [22]	95.5±0.2	98.6±0.1	**100.0±0.0**	94.4±0.3	72.9±0.2	74.9±0.3	89.4
GSDA (CVPR'20) [21]	95.7	99.1	**100**	94.8	73.5	74.9	89.7
SRDC (CVPR'20) [45]	95.7±0.2	99.2±0.1	**100.0±0.0**	95.8±0.2	76.7±0.3	77.1±0.1	90.8
RSDA (CVPR'20) [17]	96.1±0.2	99.3±0.2	**100.0±0.0**	95.8±0.3	77.4±0.8	78.9±0.3	91.1
FixBi (CVPR'21) [33]	96.1±0.2	99.3±0.2	**100.0±0.0**	95.0±0.4	**78.7±0.5**	**79.4±0.3**	91.4
CoVi (Ours)	**97.6±0.2**	**99.3±0.1**	**100.0±0.0**	**98.0±0.3**	77.5±0.3	78.4±0.3	**91.8**

Table 1: Accuracy (%) on Office-31 for unsupervised domain adaptation (ResNet-50). The best accuracy is indicated in bold, and the second-best accuracy is underlined. * Reproduced by [6].

- VisDA-C [35] is a large-scale dataset for synthetic-to-real domain adaptation across 12 categories. It contains 152,397 synthetic images for the source domain and 55,388 real-world images for the target domain.
- PACS [26] is organized into seven categories with 9,991 images in four domains: Photo (P), Art Painting (A), Cartoon (C), and Sketch (S).

4.1   Experimental Setups

Following the standard UDA protocol [14,13], we utilize labeled source data and unlabeled target data. We exploit ResNet-50 [20,19] for Office-31 and Office-Home, and ResNet-101 for VisDA-C. For multi-source domain adaptation, we use ResNet-18 in the PACS dataset. We use stochastic gradient descent (SGD) with a momentum of 0.9 in all experiments and follow the same learning rate schedule as in [13]. For the contrastive loss and label consensus loss, we follow the confidence masking policy of [33] that adaptively changes according to the sample mean and standard deviation across all mini-batches. Meanwhile, we design the *EMP-learner* by using four convolutional layers, regardless of the dataset. More detailed information is provided in the supplementary materials.

4.2   Comparison with the State-of-the-Art Methods

We validate our method compared with the state-of-the-art methods on three public benchmarks, including Office-31, Office-Home, and VisDA-C.

Office-31. In Table 1, we show the comparative performance on ResNet-50. We achieve an accuracy of 91.8%, which is 5.3% higher than the baseline MSTN [50], surpassing other state-of-the-art methods. Our method performs best in four out of six situations, *e.g.*, A→W, D→W, W→D, and A→D tasks. In particular, in A→W and A→D, although the performance improvement of the recent methods has stagnated, our method achieves a significant performance gain. We also attain better performance than the Mixup-based methods, *i.e.*, DMRL [48] and FixBi [33].

Office-Home. Table 2 demonstrates the comparison results on the Office-Home dataset based on ResNet-50. Our method achieves the highest accuracy

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.
MSTN* (Baseline) [50]	49.8	70.3	76.3	60.4	68.5	69.6	61.4	48.9	75.7	70.9	55	81.1	65.7
AADA (ECCV'20) [52]	54.0	71.3	77.5	60.8	70.8	71.2	59.1	51.8	76.9	71.0	57.4	81.8	67.0
ETD (CVPR'20) [27]	51.3	71.9	**85.7**	57.6	69.2	73.7	57.8	51.2	79.3	70.2	57.5	82.1	67.3
GSDA (CVPR'20) [21]	**61.3**	76.1	79.4	65.4	73.3	74.3	65	53.2	80	72.2	60.6	83.1	70.3
GVB-GD (CVPR'20) [10]	57	74.7	79.8	64.6	74.1	74.6	65.2	55.1	81	74.6	59.7	84.3	70.4
TCM (ICCV'21) [54]	58.6	74.4	79.6	64.5	74.0	75.1	64.6	56.2	80.9	74.6	60.7	84.7	70.7
RSDA (CVPR'20) [17]	53.2	<u>77.7</u>	<u>81.3</u>	66.4	74	76.5	<u>67.9</u>	53	<u>82</u>	75.8	57.8	85.4	70.9
SRDC (CVPR'20) [45]	52.3	76.3	81	**69.5**	76.2	<u>78</u>	**68.7**	53.8	81.7	76.3	57.1	85	71.3
MetaAlign (CVPR'21) [47]	<u>59.3</u>	76.0	80.2	65.7	74.7	75.1	65.7	56.5	81.6	74.1	61.1	85.2	71.3
FixBi (CVPR'21) [33]	58.1	77.3	80.4	67.7	<u>79.5</u>	**78.1**	65.8	<u>57.9</u>	81.7	<u>76.4</u>	62.9	**86.7**	72.7
CoVi (Ours)	58.5	**78.1**	80.0	<u>68.1</u>	**80.0**	77.0	66.4	**60.2**	82.1	**76.6**	63.6	<u>86.5</u>	**73.1**

Table 2: Accuracy (%) on Office-Home for unsupervised domain adaptation (ResNet-50). The best accuracy is indicated in bold, and the second-best accuracy is underlined. * Reproduced by [17].

Method	aero	bicycle	bus	car	horse	knife	motor	person	plant	skate	train	truck	Avg.
MSTN* (Baseline) [50]	89.3	49.5	74.3	67.6	90.1	16.6	**93.6**	70.1	86.5	40.4	83.2	18.5	65.0
DMRL (ECCV'20) [48]	-	-	-	-	-	-	-	-	-	-	-	-	75.5
TCM (ICCV'21) [54]	-	-	-	-	-	-	-	-	-	-	-	-	75.8
DWL (CVPR'21) [49]	90.7	80.2	86.1	67.6	92.4	81.5	86.8	78.1	90.6	57.1	85.6	28.7	77.1
CGDM (ICCV'21) [11]	93.4	82.7	73.2	68.4	92.9	94.5	88.7	82.1	93.4	82.5	86.8	<u>49.2</u>	82.3
STAR (CVPR'20) [29]	95	84	84.6	73	91.6	91.8	85.9	78.4	94.4	84.7	87	42.2	82.7
CAN (CVPR'19) [23]	**97**	<u>87.2</u>	82.5	74.3	**97.8**	**96.2**	90.8	80.7	<u>96.6</u>	**96.3**	87.5	**59.9**	<u>87.2</u>
FixBi (CVPR'21) [33]	96.1	**87.8**	**90.5**	**90.3**	<u>96.8</u>	<u>95.3</u>	<u>92.8</u>	**88.7**	**97.2**	<u>94.2</u>	<u>90.9</u>	25.7	<u>87.2</u>
CoVi (Ours)	<u>96.8</u>	85.6	<u>88.9</u>	<u>88.6</u>	**97.8**	93.4	91.9	<u>87.6</u>	96.0	93.8	**93.6**	48.1	**88.5**

Table 3: Accuracy (%) on VisDA-C for unsupervised domain adaptation (ResNet-101). The best accuracy is indicated in bold, and the second-best accuracy is underlined. * Reproduced by [6].

in half of the tasks and is the first to break the 73% barrier. In particular, we attain over 10% higher performance from the baseline in Cl→Pr and Pr→Cl. In addition, our method outperforms MetaAlign [47], which uses meta-learning schemes, and FixBi [33], which operates two backbone networks (*i.e.,* ResNet).

VisDA-C. In Table 3, we validate our method on a large visual domain adaptation challenge dataset with ResNet-101. Our method outperforms the state-of-the-art methods with an accuracy of 88.5%. Compared to the baseline MSTN [50], our method achieves a performance improvement of over 23%. In addition, our method shows better performance than the mixup-based DMRL [48] and FixBi [33]. We could not achieve the best accuracy across all categories due to the poor accuracy of the baseline (65.0%), yet the overall score supports the effectiveness of our method.

4.3   Ablation Studies and Discussions

Analysis of EMP. We provide visual examples of the predictions of vicinal instances using Grad-CAM [39] in Figure 4. Grad-CAM highlights class-discriminative region in an instance; hence, we can identify the most dominant label in each vicinal instance. Now we demonstrate our crucial observations based on the EMP. First, we observe that the EMP is formed differently depending on the convex combinations of the source and target instances. Second, the dominant labels are switched between the source and target labels at the
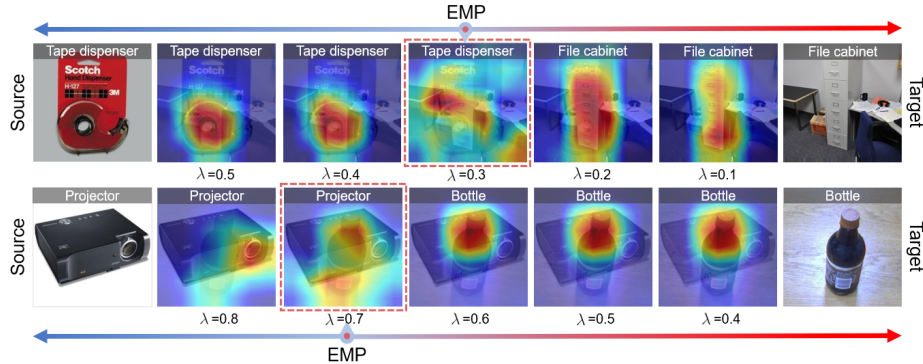
Fig. 4: **Grad-CAM visualization.** Our key observations in the vicinal space are as follows: (i) *EMPs* vary depending on the convex combination of instances. (ii) The *top-1* prediction is switched between the source and target labels (*e.g.,* Tape dispenser ↔ File cabinet) around the EMP. (iii) Grad-CAM highlights the same category as our *top-1* prediction as the most class-discriminative region.

EMP. Lastly, because the EMP is the highest entropy point, Grad-CAM fails to highlight one specific category at this point adequately. We claim that this outcome is due to the uncertainty arising from the confusion between the source and target instances. Furthermore, we discover that the source and target classes are highlighted in instances on both sides of the EMP.

Equilibrium collapse. In Figure 5, we analyze the dominance of labels between the source and target domains. Before adaptation (*i.e.*, source-only), the equilibrium of the labels is broken by the dominant-source and recessive-target domains. In this case, even if the proportion of the target instances in the mixed-up instance is more than half (*i.e.*, target-dominant instance), the top-1 predicted label is determined by the source label (*i.e.*, source-dominant label). In other words, the EMP is formed where it is biased towards the source domain. By contrast, after applying our method, we achieve equilibrium at around 50%, which is similar to the results of the supervised learning method. Consequently, our method alleviates the equilibrium collapse so that the target-dominant instance is properly predicted as a target label rather than a source label.

Analysis of the vicinal space. Our method leverages the vicinal spaces by dividing them into a contrastive space and a consensus space. In Figure 6, we observe that the *top-5* predictions of the two spaces have different characteristics. In the contrastive space, the top-2 predictions consist of the target label (*i.e.*, mobile phone) and source label (*i.e.*, backpack). In other words, the uncertainty between inter-domain categories is the most critical factor in the predictions. By contrast, in the top-2 predictions of the consensus space, the second label is not the source label but another category (*i.e.*, trash can) in the target domain that looks similar to the target label (*i.e.*, mobile phone). Hence, mitigating the intra-domain confusion of the target domain in the consensus space can be another starting point to improve performance further.
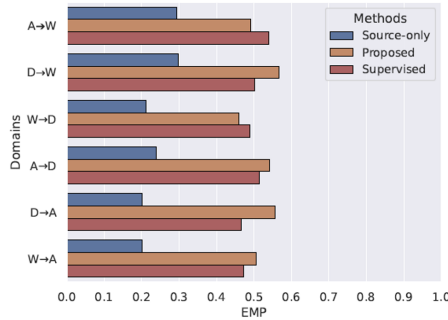
Fig. 5: **Equilibrium collapse of labels.** We compare the change of entropy maximization point according to the methods. Before adaptation, the source domain is dominant over the target domain. Contrarily, applying our method equilibrates around 50%, similar to supervised learning.
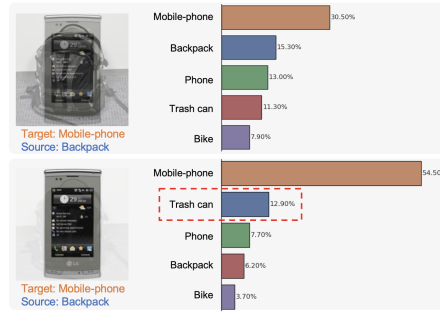

Fig. 6: **Predictions in the contrastive space *vs.* consensus space. Top:** The first and second predicted labels consist of source and target labels in the contrastive space. **Bottom:** In the consensus space, the second predicted label is not the source label but a label of another category.

Baseline	$\mathcal{R}_{emp}$	$\mathcal{R}_{ct}$	$\mathcal{R}_{cs}$	A→W	D→W	W→D	A→D	D→A	W→A	Avg.
✓				91.3	98.9	100.0	90.4	72.7	65.6	86.5
✓	✓			95.9	99.1	100.0	95.6	76.3	75.4	90.4
✓	✓	✓		97.1	99.2	100.0	97.2	76.4	76.4	91.1
✓	✓	✓	✓	**97.6**	**99.3**	**100.0**	**98.0**	**77.5**	**78.4**	**91.8**

Table 4: Ablation results (%) of investigating the effects of our components on Office-31.

Effect of the components. We conduct ablation studies to investigate the effectiveness for each component of our method in Table 4. We observe that our *EMP-Mixup* improves the accuracy by an average of 3.9% compared to the baseline [50]. In addition, our contrastive loss shows a substantial improvement in the tasks A→W and A→D. Meanwhile, in the tasks of D→A and W→A, our label-consensus loss significantly impacts the performance gain. Overall, our proposed method improves the baseline by an average of 5.3%. This experiment verifies that each component contributes positively to performance improvement.

Multi-source domain adaptation. To demonstrate the generality of our instance-wise approach, we experiment with a multi-source domain adaptation task, as shown in Table 5. Our method achieves a performance improvement of over 6% on the PACS dataset compared to the baseline MSTN [50]. In terms of the average accuracy, our method shows a significant performance improvement compared to the state-of-the-art methods. In particular, our method outperforms three out of four tasks when compared with the recent methods.

Feature visualization. We visualize the embedded features on task A→D of the Office-31 dataset using t-SNE [30] in Figure 7. Before adaptation, the source embeddings are naturally more cohesive than that of target features because only

Method	C,S,P→A	A,S,P→C	A,C,P→S	A,C,S→P	Avg.
MSTN* (Baseline) [50]	85.5	86.22	80.81	95.27	86.95
JiGen (CVPR'19) [4]	86.1	87.6	73.4	98.3	86.3
Meta-MCD (ECCV'20) [25]	87.4	86.18	78.26	97.13	87.24
CMSS (ECCV'20) [53]	88.6	90.4	82	96.9	89.5
DSON (ECCV'20) [40]	86.54	88.61	<u>86.93</u>	**99.42**	90.38
T-SVDNet (ICCV'21) [28]	<u>90.43</u>	<u>90.61</u>	85.49	98.5	<u>91.25</u>
CoVi (Ours)	**93.11**	**93.86**	**88.06**	<u>99.04</u>	**93.52**

Table 5: Accuracy (%) on PACS for multi-source unsupervised domain adaptation (ResNet-18). The best accuracy is indicated in bold, and the second-best accuracy is underlined. * Reproduced by ourselves.

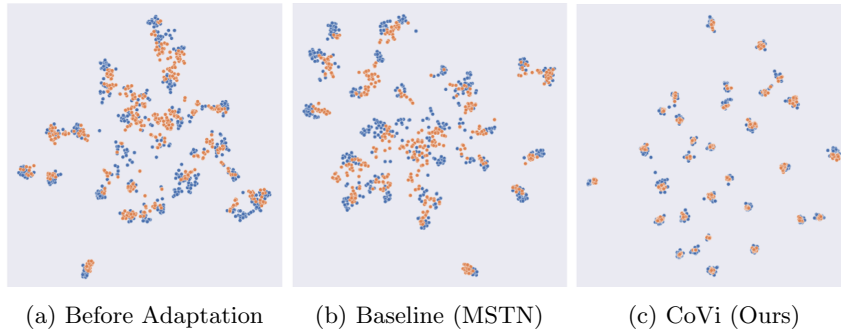(a) Before Adaptation   (b) Baseline (MSTN)   (c) CoVi (Ours)

Fig. 7: **t-SNE visualization.** Visualization of embedded features on task A→D. Blue and orange points denote the source and target domains, respectively.

source supervision is accessible. After applying the baseline (*i.e.*, MSTN [50]), we observe that the cohesion of the target features is improved but still fails to form tight clusters. By contrast, in our method, the target features construct compact clusters comparable to the source features. These results prove that our method works successfully in the unsupervised domain adaptation task.

5   Conclusions

In this study, we investigated the vicinal space between the source and target domains from the perspective of self-training. We raised the problem of the equilibrium collapse of labels and proposed three novel approaches. Our EMP-Mixup efficiently minimized the worst-case risk in the vicinal space. In addition, we reduce inter-domain and intra-domain confusions by dividing the vicinal space into contrastive and consensus space. The competitiveness of our approach suggests that self-predictions in vicinal space can play an important role in solving the UDA problem.

References

1. Berthelot, D., Carlini, N., Cubuk, E.D., Kurakin, A., Sohn, K., Zhang, H., Raffel, C.: Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. arXiv preprint arXiv:1911.09785 (2019) 3, 4
2. Berthelot, D., Carlini, N., Goodfellow, I., Papernot, N., Oliver, A., Raffel, C.: Mixmatch: A holistic approach to semi-supervised learning. arXiv preprint arXiv:1905.02249 (2019) 4
3. Bousmalis, K., Trigeorgis, G., Silberman, N., Krishnan, D., Erhan, D.: Domain separation networks. Advances in neural information processing systems **29**, 343–351 (2016) 2
4. Carlucci, F.M., D'Innocente, A., Bucci, S., Caputo, B., Tommasi, T.: Domain generalization by solving jigsaw puzzles. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2229–2238 (2019) 14
5. Chang, W.L., Wang, H.P., Peng, W.H., Chiu, W.C.: All about structure: Adapting structural information across domains for boosting semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1900–1909 (2019) 2
6. Chang, W.G., You, T., Seo, S., Kwak, S., Han, B.: Domain-specific batch normalization for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7354–7362 (2019) 4, 10, 11
7. Chapelle, O., Weston, J., Bottou, L., Vapnik, V.: Vicinal risk minimization. Advances in neural information processing systems pp. 416–422 (2001) 5
8. Cubuk, E.D., Zoph, B., Mane, D., Vasudevan, V., Le, Q.V.: Autoaugment: Learning augmentation policies from data. arXiv preprint arXiv:1805.09501 (2018) 4
9. Cubuk, E.D., Zoph, B., Shlens, J., Le, Q.V.: Randaugment: Practical automated data augmentation with a reduced search space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 702–703 (2020) 4
10. Cui, S., Wang, S., Zhuo, J., Su, C., Huang, Q., Tian, Q.: Gradually vanishing bridge for adversarial domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12455–12464 (2020) 2, 11
11. Du, Z., Li, J., Su, H., Zhu, L., Lu, K.: Cross-domain gradient discrepancy minimization for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3937–3946 (2021) 11
12. Fan, K.: Minimax theorems. Proceedings of the National Academy of Sciences of the United States of America **39**(1), 42 (1953) 2, 6
13. Ganin, Y., Lempitsky, V.: Unsupervised domain adaptation by backpropagation. In: International conference on machine learning. pp. 1180–1189. PMLR (2015) 3, 10
14. 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) 10
15. Gong, R., Li, W., Chen, Y., Gool, L.V.: Dlow: Domain flow for adaptation and generalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2477–2486 (2019) 2
16. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014) 2, 6

17. Gu, X., Sun, J., Xu, Z.: Spherical space domain adaptation with robust pseudo-label loss. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9101–9110 (2020) 2, 4, 10, 11

18. Guo, H., Mao, Y., Zhang, R.: Mixup as locally linear out-of-manifold regularization. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 3714–3722 (2019) 4

19. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016) 10

20. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: European conference on computer vision. pp. 630–645. Springer (2016) 10

21. Hu, L., Kan, M., Shan, S., Chen, X.: Unsupervised domain adaptation with hierarchical gradient synchronization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4043–4052 (2020) 10, 11

22. Jin, Y., Wang, X., Long, M., Wang, J.: Minimum class confusion for versatile domain adaptation. In: European Conference on Computer Vision. pp. 464–480. Springer (2020) 10

23. Kang, G., Jiang, L., Yang, Y., Hauptmann, A.G.: Contrastive adaptation network for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 4893–4902 (2019) 4, 11

24. Lee, D.H., et al.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, ICML. p. 896 (2013) 2

25. Li, D., Hospedales, T.: Online meta-learning for multi-source and semi-supervised domain adaptation. In: European Conference on Computer Vision. pp. 382–403. Springer (2020) 14

26. Li, D., Yang, Y., Song, Y.Z., Hospedales, T.M.: Deeper, broader and artier domain generalization. In: Proceedings of the IEEE international conference on computer vision. pp. 5542–5550 (2017) 3, 10

27. Li, M., Zhai, Y.M., Luo, Y.W., Ge, P.F., Ren, C.X.: Enhanced transport distance for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13936–13944 (2020) 11

28. Li, R., Jia, X., He, J., Chen, S., Hu, Q.: T-svdnet: Exploring high-order prototypical correlations for multi-source domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9991–10000 (2021) 14

29. Lu, Z., Yang, Y., Zhu, X., Liu, C., Song, Y.Z., Xiang, T.: Stochastic classifiers for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9111–9120 (2020) 11

30. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008) 13

31. Mai, Z., Hu, G., Chen, D., Shen, F., Shen, H.T.: Metamixup: Learning adaptive interpolation policy of mixup with metalearning. IEEE Transactions on Neural Networks and Learning Systems (2021) 4

32. Miyato, T., Maeda, S.i., Koyama, M., Ishii, S.: Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE transactions on pattern analysis and machine intelligence **41**(8), 1979–1993 (2018) 6

33. Na, J., Jung, H., Chang, H.J., Hwang, W.: Fixbi: Bridging domain spaces for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1094–1103 (2021) 1, 2, 4, 10, 11

34. Pan, F., Shin, I., Rameau, F., Lee, S., Kweon, I.S.: Unsupervised intra-domain adaptation for semantic segmentation through self-supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3764–3773 (2020) 4

35. Peng, X., Usman, B., Kaushik, N., Hoffman, J., Wang, D., Saenko, K.: Visda: The visual domain adaptation challenge. arXiv preprint arXiv:1710.06924 (2017) 3, 10

36. Pham, H., Dai, Z., Xie, Q., Le, Q.V.: Meta pseudo labels. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11557–11568 (2021) 2

37. Saenko, K., Kulis, B., Fritz, M., Darrell, T.: Adapting visual category models to new domains. In: European conference on computer vision. pp. 213–226. Springer (2010) 3, 9

38. Saito, K., Ushiku, Y., Harada, T.: Asymmetric tri-training for unsupervised domain adaptation. In: International Conference on Machine Learning. pp. 2988–2997. PMLR (2017) 2

39. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017) 11

40. Seo, S., Suh, Y., Kim, D., Kim, G., Han, J., Han, B.: Learning to optimize domain specific normalization for domain generalization. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXII 16. pp. 68–83. Springer (2020) 14

41. Sharma, A., Kalluri, T., Chandraker, M.: Instance level affinity-based transfer for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5361–5371 (2021) 10

42. Shimodaira, H.: Improving predictive inference under covariate shift by weighting the log-likelihood function. Journal of statistical planning and inference **90**(2), 227–244 (2000) 1

43. Shin, I., Woo, S., Pan, F., Kweon, I.S.: Two-phase pseudo label densification for self-training based domain adaptation. In: European conference on computer vision. pp. 532–548. Springer (2020) 2

44. Sohn, K., Berthelot, D., Li, C.L., Zhang, Z., Carlini, N., Cubuk, E.D., Kurakin, A., Zhang, H., Raffel, C.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. arXiv preprint arXiv:2001.07685 (2020) 3, 4

45. Tang, H., Chen, K., Jia, K.: Unsupervised domain adaptation via structurally regularized deep clustering. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 8725–8735 (2020) 10, 11

46. Venkateswara, H., Eusebio, J., Chakraborty, S., Panchanathan, S.: Deep hashing network for unsupervised domain adaptation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5018–5027 (2017) 3, 9

47. Wei, G., Lan, C., Zeng, W., Chen, Z.: Metaalign: Coordinating domain alignment and classification for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 16643–16653 (2021) 11

48. Wu, Y., Inkpen, D., El-Roby, A.: Dual mixup regularized learning for adversarial domain adaptation. In: European Conference on Computer Vision. pp. 540–555. Springer (2020) 2, 4, 10, 11

49. Xiao, N., Zhang, L.: Dynamic weighted learning for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15242–15251 (2021) 1, 10, 11

18   J. Na et al.

50. Xie, S., Zheng, Z., Chen, L., Chen, C.: Learning semantic representations for unsupervised domain adaptation. In: International conference on machine learning. pp. 5423–5432. PMLR (2018) 3, 4, 10, 11, 13, 14
51. Xu, M., Zhang, J., Ni, B., Li, T., Wang, C., Tian, Q., Zhang, W.: Adversarial domain adaptation with domain mixup. In: Proceedings of the AAAI Conference on Artificial Intelligence. pp. 6502–6509 (2020) 2, 4
52. Yang, J., Zou, H., Zhou, Y., Zeng, Z., Xie, L.: Mind the discriminability: Asymmetric adversarial domain adaptation. In: European Conference on Computer Vision. pp. 589–606. Springer (2020) 11
53. Yang, L., Balaji, Y., Lim, S.N., Shrivastava, A.: Curriculum manager for source selection in multi-source domain adaptation. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIV 16. pp. 608–624. Springer (2020) 14
54. Yue, Z., Sun, Q., Hua, X.S., Zhang, H.: Transporting causal mechanisms for unsupervised domain adaptation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8599–8608 (2021) 1, 11
55. Zhang, H., Cisse, M., Dauphin, Y.N., Lopez-Paz, D.: mixup: Beyond empirical risk minimization. arXiv preprint arXiv:1710.09412 (2017) 2, 4, 5
56. Zhu, J., Shi, L., Yan, J., Zha, H.: Automix: Mixup networks for sample interpolation via cooperative barycenter learning. In: European Conference on Computer Vision. pp. 633–649. Springer (2020) 4