Towards Effective Instance Discrimination Contrastive Loss for Unsupervised Domain Adaptation

Yixin Zhang[1,2] Zilei Wang[*2] Junjie Li[2] Jiafan Zhuang[3] Zihan Lin[2]
[1] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center
[2] University of Science and Technology of China
[3] Shantou University
{zhyx12, zlwang}@ustc.edu.cn, {hnljj, myustc}@mail.ustc.edu.cn, jfzhuang@stu.edu.cn

Abstract

*Domain adaptation (DA) aims to transfer knowledge from a label-rich source domain to a related but label-scarce target domain. Recently, increasing research has focused on exploring data structure of the target domain. In light of the recent success of **I**nstance **D**iscrimination **Co**ntrastive (IDCo) loss in self-supervised learning, we try directly applying it to domain adaptation tasks. However, the improvement is very limited, which motivates us to re-think its underlying limitations for domain adaptation tasks. An intuitive limitation is that a pair of samples belonging to the same class could be treated as negatives. Here we argue that using low-confidence samples to construct positive and negative pairs can alleviate this issue and is more suitable for IDCo loss. Another limitation is that IDCo loss cannot capture enough semantic information. We address this by introducing domain-invariant and accurate semantic information from classifier weights and input data. Specifically, we propose a class relationship enhanced features. It uses probability weighted class prototpyes as the input features of IDCo loss, which can implicitly transfer the domain-invariant class relationship. We further propose a target-dominated cross-domain mixup that can incorporate accurate semantic information from the source domain. We evaluate the proposed method in unsupervised DA and other DA settings, and extensive experimental results reveal that our method can make IDCo loss more effective and achieve state-of-the-art performance.*[1]

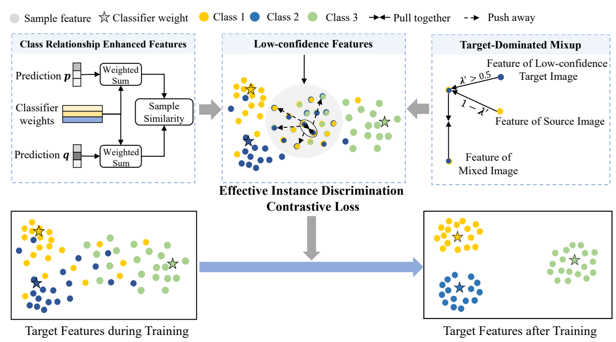

Figure 1. Illustration of our proposed method. The bottom left represents the target-domain feature distribution during training. The bottom right is the feature distribution obtained by our method. The upper part is our proposed effective instance discrimination contrastive loss, which contains three key designs. Best viewed in color.

1. Introduction

Deep neural networks have shown considerable effectiveness in a variety of machine learning challenges [29, 5, 56]. However, the impressive performance gain heavily relies on massive well-labeled training data. Additionally, manually annotating sufficient training data is often time and expense prohibitive in reality. Besides, another disadvantage of traditional deep learning is its inability to generalize to new datasets due to the domain shift problem [2, 1]. Domain Adaptation (DA) addresses this issue by utilizing the knowledge of a label-rich source domain to assist the learning in a related but label-scarce target domain.

The most popular way to deal with domain shift is to learn domain-invariant representations. We can roughly classify these DA approaches as either discrepancy metric-based methods [45, 83] or adversarial-based methods [72, 42, 9]. Recently, there have been more methods exploring the inherent structures of unlabeled target domains, such as self-training through pseudo labels [15, 101, 27, 43] and aligning prototypes across domains [25, 95]. The majority of them rely on trustworthy samples selected by some criteria, such as probability [52, 3] and sample ratio [101]. This can create reliable pseudo labels or prototypes, leading to better discriminability on the target domain. However, this

*Corresponding author
[1]Code is available at https://github.com/zhyx12/EIDCo

would lead to a suboptimal transferability since the trustworthy samples are more biased to source domain samples and meanwhile less confident samples are not well learned.

Recently, **I**nstance **D**iscrimination **Co**ntrastive (*IDCo*) loss [19, 6] has achieved great success in self-supervised learning. It regards different views of the same image as positive pairs and pulls them togher. All other images are negative samples and are pushed away from the query sample. Inspired by this, a natural idea is to introduce *IDCo* loss into the unlabeled target domain, considering no class labels are needed. However, when directly applying *IDCo* loss in feature space, we observed that only a slight improvement was gained. We conjecture that it is caused by two main factors. Firstly, there exists category collision in *IDCo* loss [91, 26]. That is, two instances, even belonging to the same category, are considered negative pairs if they originate from different samples, and their similarity will be reduced. Secondly, with little category priors, the traditional *IDCo* loss learn rich low-level features without encoding enough high-level semantic informantion [61, 71, 21]. This is suboptimal for many visual recognition tasks that require discriminative semantic features.

Existing work improves *IDCo* loss by selecting more informative positive and negative samples, *i.e.*, integrating category information [21] or resorting to the nearest neighbors [14, 90]. Although the limitations can be partially alleviated, the selection of informative samples are complicated. Differently, we focus on exploring a *pure* and *effective IDCo* loss for domain adaptation tasks.

For the first limitation, we argue that category collision cannot be completely avoided due to lack of accurate labels. However, considering the *push away* process between the query sample and negative samples, the contributions of different negative samples are not equal [74]. The closer negative sample contributes more to the *push away* process. Based on this, we find that low-confidence samples are more suitable for *IDCo* loss. Specifically, if we draw both positive and negative samples from low-confidence ones as in Figure 1 upper middle, the closer negative samples will more likely belong to different classes with the query sample. With this simple design, category collision can be greatly alleviated. More importantly, low-confidence sample based *IDCo* loss is complementary to existing self-training or category contrastive methods, which rely more on high-confidence samples.

For the second limitation, it is necessary to explicitly involve semantic information into *IDCo* loss. We solve this by introducing domain-invariant and accurate semantic information using classifier weights and input data. On the one hand, the classifier weights can be regarded as class prototypes. Instead of using original features, we propose class relationship enhanced features where the classifier prototypes is weighted by predicted probability. Through this re-represented features, the domain-invariant class relationship can be implicitly embedded when computing the similarity of two samples. Specifically, the class relationship is represented by cosine similarities matrix $\boldsymbol{A}$ of the classifier weights, and the similarity of two samples can be represented as the sum of $\boldsymbol{A}$ weighted outer product of two probabilities. In this way, the class relationship can be better maintained. On the other hand, the source images contain accurate semantic information. We propose to combine Mixup [93] with *IDCo* loss. The model is encouraged to behave linearly across the source and target domains during the instance discrimination process. In such a case, it is still crucial to guarantee the low-confidence of mixed samples according to the above analysis. To this end, we propose a target-dominated mixup where the low-confidence target samples are given a higher weight than the source samples. By combining all three designs, the *IDCo* loss can work well, and features of low-confidence target samples are better learned.

Our contributions are summarized as follows:

- We propose a *pure* and *effective IDCo* loss in domain adaptation for image classification. Here two main limitations are considered and can be greatly alleviated.
- We propose to use only low-confidence samples in *IDCo* loss to alleviate category collision. Furthermore, we propose class relationship enhanced features and target-dominated cross-domain mixup to encode domain-invariant and accurate semantic information.
- We conducted extensive experiments on multiple DA benchmarks, and the results reveal that our proposed method achieves the state-of-the-art performance.

2. Related Work

2.1. Unsupervised Domain Adaptation

Unsupervised Domain Adaptation (UDA) [2, 1] is a technique for generalizing a model from a labeled source domain to an unlabeled target domain. The mainstream approaches are to learn domain-invariant representations, and they can be classified into domain discrepancy based methods [45, 83, 66, 67, 53] and domain adversarial training based methods [17, 72, 97, 42, 76, 39].

To increase discriminability, more recent methods attempt to investigate data structure in unlabeled target domain. Self-training as a typical approach generates target pseudo labels [15, 101, 48, 51, 27, 30, 43]. Another line is to construct prototypes [52, 3, 98, 95] or cluster centers [25, 10, 69] across domains and then perform class-wise alignment. Most of these approaches use probability threshold or sample ratio to choose the trustworthy samples and neglect other less trustworthy samples. To mitigate the harmful effect of noisy labels, it is reasonable to utilize only reliable samples. However, less trustworthy samples are also important since they may reveal the complete structure of the target data.

We turn to instance discrimination contrastive loss to explore the data distribution in the target domain. It should be noted that most approaches [25, 91, 21] in domain adaptation use contrastive loss on the basis of classes (not instances) with a selection strategy of reliable samples. Only a few methods [26, 57, 63] use the instance discrimination contrastive loss. CDS [26] employs contrastive learning in a pre-train step before proceeding to a domain alignment stage. CLDA [63] suggests that the classifier can be used as a contrastive projection head [6]. CoMix [57] conducts temporal contrastive self-supervised learning over the graph representations. In contrast to previous efforts, we focus on the potential limitations of the *IDCo* loss in DA tasks, and then propose three key designs to make it more effective.

2.2. Related Techniques

Instance discrimination contrastive learning has shown remarkable advantages in self-supervised learning [19, 6, 14]. The contrastive loss measures the similarity of representation pairs and attempts to distinguish between positive and negative pairs. MoCo [19] maintains a queue of previously processed embeddings as a negative memory bank. SimCLR [6] shows that large batch size and strong data augmentations have a comparable performance to the memory-based approaches. Here we adopt a similar architecture to MoCo [19] to perform contrastive learning. Differently, we carefully design the input features to more concern the low-confidence samples, and propose a novel class relationship enhanced features and target-dominated cross-domain mixup to encode semantic information.

Mixup [93] provide effective data augmentation strategies for supervised and semi-supervised learning. In domain adaptation, the domain-level mixup [80, 79] and category-level mixup [84, 49, 89] are used to learn domain-invarient features. In self-supervised learning, recent work has leveraged the idea of image space mixtures [62, 32] and embedding space mixtures [24, 99] to generate more valuable positive or negative samples. We propose a target-dominated cross-domain mixup which is more compatible with *IDCo* loss in DA tasks.

3. Preliminary

Unsupervised domain adaptation (UDA) for classification aims to train a model on labeled source domain $\mathcal{D}_s = \{(\boldsymbol{x}_s^i, y_s^i)\}_{i=1}^{n_s}$ and unlabeled target domain $\mathcal{D}_t = \{\boldsymbol{x}_t^j\}_{j=1}^{n_t}$ to obtain high accuracy on a target domain test set. The data in source and target domains are drawn from the joint distributions $P(\boldsymbol{x}_s, y_s)$ and $Q(\boldsymbol{x}_t, y_t)$ with $P \neq Q$.

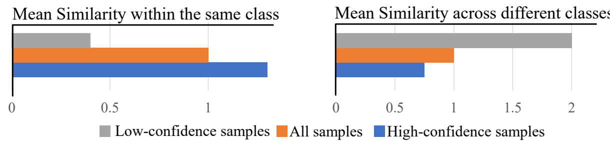

Figure 2. Motivation for exploiting low-confidence samples: the mean similarity within the same class (left) and across classes (right). In each figure, all three similarities are scaled, making the value for all samples similarity equal to 1 for better comparison. Best viewed in color.

The framework is shown in Figure 3. We first split the network into a feature extractor F and a classifier C, and then adopt a teacher-student framework. The teacher model ($\tilde{F}$ and $\tilde{C}$) is continuously updated by exponential moving average (EMA) [70] of the student model (F and C). For a target sample, we send its weakly augmented view x_t^0 to the teacher model and obtain its probability $\tilde{\boldsymbol{p}}_t^0$. Then we compare the maximum value of $\tilde{\boldsymbol{p}}_t^0$ with the predefined probability threshold τ. If $\max(\tilde{\boldsymbol{p}}_t^0) > \tau$, it is high confidence. Otherwise, it is low-confidence. We adopt FixMatch [65] for high-confidence ones. In this work, we focus on low-confidence target samples.

Before introducing our method, we first review the traditional **I**nstance **D**iscrimination **Co**ntrastive (*IDCo*) loss used in self-supervised learning. Given an image $\boldsymbol{x}_t \in \mathcal{D}_t$, we can obtain two strongly augmented views $\boldsymbol{x}_t^1$, $\boldsymbol{x}_t^2$ as the query and the key image. Then ℓ_2-normalized features can be produced by $\boldsymbol{f}_i = \ell_2(F(\boldsymbol{x}_t^1)), \tilde{\boldsymbol{f}}_t = \ell_2(\tilde{F}(\boldsymbol{x}_t^2))$. The naive contrastive loss without other designs (*e.g.*, projection head) can be presented as follows:

$$h(\boldsymbol{f}_t, \tilde{\boldsymbol{f}}_t) = \exp(\boldsymbol{f}_t^T \tilde{\boldsymbol{f}}_t / T_{co}),$$
$$\ell_{idco} = -\log \frac{h(\boldsymbol{f}_t, \tilde{\boldsymbol{f}}_t)}{h(\boldsymbol{f}_t, \tilde{\boldsymbol{f}}_t) + \sum\limits_{\tilde{\boldsymbol{f}}_- \in M} h(\boldsymbol{f}_t, \tilde{\boldsymbol{f}}_-)}, \tag{1}$$

where T_{co} is the temperature hyperparameter, and we use h to denote the exponential of scaled cosine similarity. M is the memory bank [19] that stores the features processed by the teacher feature extractor. Intuitively, this loss pulls the $\boldsymbol{f}_t$ close to $\tilde{\boldsymbol{f}}_t$, and pushes $\boldsymbol{f}_t$ away from $\tilde{\boldsymbol{f}}_-$.

In this work, we propose to use *IDCo* loss to learn the unlabeled target data. We tried directly applying it in the feature space (even with an extra projection head [6]), but got limited improvement ($< 0.5\%$), which motivates us to rethink the potential limitations of *IDCo* loss for domain adaptation tasks.

4. Our Method

4.1. Necessity of Low-confidence Samples

An intuitive limitation is that a pair of samples belonging to the same class could be treated as negatives. More

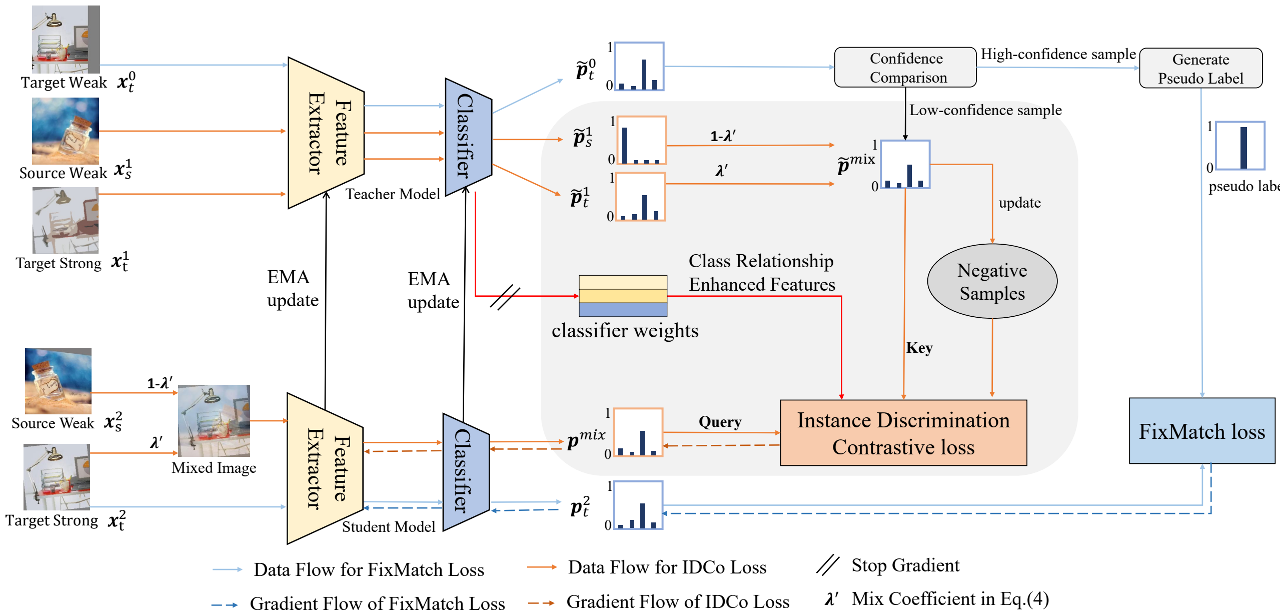

Figure 3. The framework of our proposed method. Here a teacher-student framework is adopted. For the high-confidence target samples, we use FixMatch loss. For the low-confidence ones, we propose an **E**ffective **I**nstance **D**iscrimination **Co**ntrastive (**EIDCo**) loss (*i.e.* shaded part).

specifically, for a balanced dataset D, there are on average $\frac{|D|}{K}$ samples that belong to the same class as the query image, but are regarded as negative samples (K is the number of classes). This cannot be changed as long as M is constructed randomly. Inspired by previous work [21], we delve into the average cosine similarities of features within the same class and across classes. Figure 2 depicts our findings under *Office-Home* Art→Clipart. Here, the definition of high- and low-confidence sample follows Sec. 3. It can be observed that for the low-confidence samples, the average similarity within the same class is smaller (*i.e.* gray row in Figure 2 left) and that across classes is larger (*i.e.* gray row in Figure 2 right). When considering the *push away* process between the negative sample $\tilde{\boldsymbol{f}}_-$ and the query sample $\boldsymbol{f}_t$, the distance between $\tilde{\boldsymbol{f}}_-$ and $\boldsymbol{f}_t$ determines the contribution of $\tilde{\boldsymbol{f}}_-$ to push $\boldsymbol{f}_t$ away [74]. Specifically, the closer negative sample $\tilde{\boldsymbol{f}}_-$ is to $\boldsymbol{f}_t$, the more contributions it would make to push $\boldsymbol{f}_t$ away.

For low-confidence samples, the similarity within the same class is still small, while the similarity across different classes is relatively larger. If we construct $\boldsymbol{f}_t$ and $\tilde{\boldsymbol{f}}_-$ from low-confidence samples, the closer negative samples around $\boldsymbol{f}_t$ are more likely to belong to different classes. As a result, the *IDCo* loss will focus more on pushing $\boldsymbol{f}_t$ away from samples of different classes. This indicates the low-confidence samples are more suitable for *IDCo* loss.

4.2. Class Relationship Enhanced Features

Another limitation of the *IDCo* loss for DA tasks is that it does not encode enough semantic information, which may result in inconsistency with task-specific discriminability. To tackle this issue, we propose to involve classifier weights in contrastive learning and introduce a domain-invariant semantic prior. Each classifier weight can be regarded as a class prototype. Although the classifier is somewhat source biased due to the supervision of source samples or high-confidence target samples [100], the semantic relationship between different classes is maintained across domains [12]. Given the normalized classifier weights $\boldsymbol{W} = [\boldsymbol{w}_1, \boldsymbol{w}_2, ...\boldsymbol{w}_K]^T \in \mathbb{R}^{K\times D}$ where D is the feature dimension, and the probabilities $\boldsymbol{p} \in \mathbb{R}^{K\times 1}$ of a low-confidence target domain sample, we propose a novel **C**lass **R**elationship enhanced **F**eatures (CRF) which can be presented as follows:

$$\boldsymbol{f}_{cr}(\boldsymbol{p}) = \sum_{i=1}^{K} p_i \boldsymbol{w}_i = (\boldsymbol{p}^T \boldsymbol{W})^T \in \mathbb{R}^{D\times 1} \tag{2}$$

Based on this re-represented feature, the $h(\cdot)$ in Eq. 1 which uses cosine similarity of features can be replaced by the following $h_{cr}(\cdot)$:

$$\begin{aligned} s_{cr}(\boldsymbol{p},\boldsymbol{q}) = \boldsymbol{f}_{cr}(\boldsymbol{p})^T \boldsymbol{f}_{cr}(\boldsymbol{q}) = (\boldsymbol{p}^T \boldsymbol{W})(\boldsymbol{q}^T \boldsymbol{W})^T, \\ h_{cr}(\boldsymbol{p},\boldsymbol{q}) = \exp(s_{cr}(\boldsymbol{p},\boldsymbol{q})/T_{co}), \end{aligned} \tag{3}$$

where s_{cr} represents the similarity of two samples (*i.e.* $\boldsymbol{p}$ and $\boldsymbol{q}$), T_{co} is the same temperature as in Eq. 1.

In the above process, the class relationship is implicitly embedded. Specifically, we use the consine similarity of classifier prototypes to represent the class relationship. Here, the class similarity matrix $\boldsymbol{A}$ can be obtained by

$\boldsymbol{A} = \boldsymbol{W}(\boldsymbol{W})^T \in \mathbb{R}^{K \times K}$, and the proposed similarity s_{cr} can be further presented as follows:

$$\begin{aligned} s_{cr}(\boldsymbol{p}, \boldsymbol{q}) &= \sum_{i,j=1}^{K} (p_i \boldsymbol{w}_i)^T (q_j \boldsymbol{w}_j) = \sum_{i,j=1}^{K} A_{i,j} \times p_i \times q_j \\ &= \sum_{i=1}^{K} p_i \times q_i + \sum_{i=1}^{K} \sum_{j=1, j \neq i}^{K} A_{i,j} \times p_i \times q_j \end{aligned} \tag{4}$$

In the above equation, s_{cr} is split into two terms. The former is the traditional intra-class similarity which is inspired by PCL [34]. The latter can be viewed as inter-class similarity which contains probability products of all different classes. The nondiagonal elements in $\boldsymbol{A}$ is further embedded into it as coefficients. By using the intra-class term alone, the classifier weight can be optimized by *IDCo* loss, thus can incorporate some semantic information [34]. However, intra-class term lacks the true relationship of samples in the feature space. Since $A_{i,j}$ reflects the true similarity of different class i and class j, adding the inter-class term can better express the relationship of two samples. This is important since our low-confidence sample based *IDCo* loss relies on different distances between negative samples and the query sample during *push away* process. Using intra-class similarity alone will weaken the difference of distances between negative samples and the query sample, and thus is harmful in alleviating category collision. Adding inter-class similarity will integrate class relationship prior and represent the sample relationship more accurately.

Some previous domain adaptation methods also consider class relationship. CAiDA [12] focus on multi-source-free domain adaptation, and propose to preserve the consistency of inter-class relationship through aligning soft label distributions of different domains. Some methods (*e.g.* MCC [23] and BCDM [37]) propose to utilize the inter-class term, but their primary objective is to suppress it and produce more confident predictions. It is implemented by directly minimizing the inter-class term [23] or resorting to adversarial training between two classifiers [37]. Different from them, we aim to accurately express the similarity of two samples by incorporating class relationship prior.

4.3. Target-Dominated Cross-Domain Mixup

The above proposed sample similarity can incorporate domain-invariant class relationship. Here we further introduce more accurate semantic information (*i.e.* labeled source image) at the input level. Different from previous methods which use mixup in either domain adversarial learning [80] or pseudo-label based self-training [49], we propose to introduce mixup in *IDCo* loss as shown in Figure 3. According to previous subsection 4.1, the low-confidence characteristic of input samples should be maintained. Thus, we propose a target-dominated strategy where target samples are given higher weights. Specifically, given a low-confidence target sample $\boldsymbol{x}_t \in \mathcal{D}_t$, we randomly select a source sample $\boldsymbol{x}_s \in \mathcal{D}_s$. Then we generate two views for each sample, *i.e.*, $\boldsymbol{x}_t^1, \boldsymbol{x}_t^2$ for $\boldsymbol{x}_t$ and $\boldsymbol{x}_s^1$ and $\boldsymbol{x}_s^2$ for $\boldsymbol{x}_s$. The image mix of $\boldsymbol{x}_t^2$ and $\boldsymbol{x}_s^2$ can be presented as

$$\begin{aligned} \lambda' = \max(\lambda, 1-\lambda), \lambda \sim \texttt{Beta}(\alpha, \alpha), \\ \boldsymbol{x}_t^{mix} = \lambda' \boldsymbol{x}_t^2 + (1-\lambda')\boldsymbol{x}_s^2, \end{aligned} \tag{5}$$

where α is the parameter controlling the shape of Beta distribution. In our experiments, we set $\alpha = 1.0$.

In original Mixup [93], the label is mixture of two labels. When combining mixup with *IDCo* loss, the key should be mixture of two keys as shown in Figure 3. Specifically, the *query* is student model prediction (*i.e.* $\boldsymbol{p}_t^{mix}$) of the mixed image $\boldsymbol{x}_t^{mix}$. The *key* is the mixture (*i.e.* $\tilde{\boldsymbol{p}}_t^{mix} = \lambda' \tilde{\boldsymbol{p}}_t^1 + (1-\lambda')\tilde{\boldsymbol{p}}_s^1$) of teacher model predictions (*i.e.* $\tilde{\boldsymbol{p}}_t^1, \tilde{\boldsymbol{p}}_s^1$) of images $\boldsymbol{x}_t^1$ and $\boldsymbol{x}_s^1$.

By combining all three designs, we can construct an **E**ffective **I**nstance **D**iscrimination **Co**ntrastive (**EIDCo**) loss, which represented as follows:

$$\ell_{eidco} = -\log \frac{h_{cr}(\boldsymbol{p}_t^{mix}, \tilde{\boldsymbol{p}}_t^{mix})}{\sum\limits_{\tilde{\boldsymbol{p}}_+ \in P_+} h_{cr}(\boldsymbol{p}_t^{mix}, \tilde{\boldsymbol{p}}_+) + \sum\limits_{\tilde{\boldsymbol{p}}_- \in M} h_{cr}(\boldsymbol{p}_t^{mix}, \tilde{\boldsymbol{p}}_-)}, \tag{6}$$

where $h_{cr}(\cdot)$ is defined in Eq. 3, P_+ contains $\{\tilde{\boldsymbol{p}}_t^1, \tilde{\boldsymbol{p}}_s^1\}$, M stores previous keys $\tilde{\boldsymbol{p}}_t^{mix}$ produced by teacher model. This loss can be regarded as the log form of a ($|M| + 2$)-way softmax-based classifier that tries to classify $\boldsymbol{p}_t^{mix}$ as $\tilde{\boldsymbol{p}}_t^{mix}$.

4.4. Overall Training Objective

Our proposed method focus on the learning of low-confidence target domain samples through *IDCo* loss. It cannot be used alone since the discriminativeness of target domain features can not be guaranteed. Adding existing well-performed baselines, including domain adversarial methods (*i.e.* GVB [9], ToAlign [78], and Baseline-B in SSRT [68]) or entropy adversarial method (*i.e.* MME [58]) is feasible, since their training uses all target samples and can generate high-confidence predictions. Here, we further add the simple yet effective FixMatch [65] loss to existing baseline and get a stronger one. As a result, the overall loss is presented as

$$\ell_{final} = \ell_{baseline} + \ell_{fixmatch} + \lambda_{co}\ell_{eidco}, \tag{7}$$

where $\ell_{baseline}$ represents the loss in the baseline method, $\ell_{fixmatch}$ is the same as FixMatch [65], λ_{co} is the trade-off hyperparameter.

5. Experiments

We evaluate the effectiveness of our method under three DA settings, *i.e.*, single source unsupervised domain adaptation (SUDA), semi-supervised domain adapta-

Table 1. Accuracy (%) of different UDA methods on *Office-Home* with ResNet-50 (R-50) and ViT-B (T).

Net	Method	A→C	A→P	A→R	C→A	C→P	C→R	P→A	P→C	P→R	R→A	R→C	R→P	Acc
R-50	Source-Only	34.9	50.0	58.0	37.4	41.9	46.2	38.5	31.2	60.4	53.9	41.2	59.9	46.1
	HDAN (NeurIPS'20) [22]	56.8	75.2	79.8	65.1	73.9	75.2	66.3	56.7	81.8	75.4	59.7	84.7	70.9
	PRONOUN (TIP'21) [20]	57.6	75.0	78.4	64.9	74.0	74.8	66.6	58.2	80.4	74.3	60.4	84.3	70.7
	TSA (CVPR'21) [38]	53.6	75.1	78.3	64.4	73.7	72.5	62.3	49.4	77.5	72.2	58.8	82.1	68.3
	CKB-MMD (CVPR'21) [46]	54.2	74.1	77.5	64.6	72.2	71.0	64.5	53.4	78.7	72.6	58.4	82.8	68.7
	FixBi (CVPR'21) [49]	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
	ATDOC (CVPR'21) [41]	60.2	77.8	82.2	68.5	78.6	77.9	68.4	58.4	83.1	74.8	61.5	**87.2**	73.2
	MetaAlign (CVPR'21) [77]	59.3	76.0	80.2	65.7	74.7	75.1	65.7	56.5	81.6	74.1	61.1	85.2	71.3
	SCDA (ICCV'21) [39]	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
	TCM (ICCV'21) [92]	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
	FGDA (ICCV'21) [18]	52.3	77.0	78.2	64.6	75.5	73.7	64.0	49.5	80.7	70.1	52.3	81.6	68.3
	CST (NeurIPS'21) [43]	59.0	79.6	**83.4**	68.4	77.1	76.7	68.9	56.4	83.0	75.3	62.2	85.1	73.0
	MCC+NWD (CVPR'22) [4]	58.1	79.6	83.7	67.7	77.9	78.7	66.8	56.0	81.9	73.9	60.9	86.1	72.6
	GVB (CVPR'20) [9]	57.0	74.7	79.8	64.6	74.1	74.6	65.2	55.1	81.0	74.6	59.7	84.3	70.4
	+ **EIDCO**	58.1	76.8	80.0	65.1	75.4	74.4	65.6	58.3	82.2	75.0	62.8	84.7	71.5
	+ FixMatch	61.0	77.6	80.4	65.6	76.8	74.5	66.5	60.0	83.3	76.7	64.8	85.1	72.7
	+ FixMatch + **EIDCo**	**64.4**	**81.1**	81.6	68.5	78.9	78.8	69.1	59.9	**87.0**	77.3	**67.7**	86.7	75.1
	ToAlign (NeurIPS'21) [78]	57.9	76.9	80.8	66.7	75.6	77.0	67.8	57.0	82.5	75.1	60.0	84.9	72.0
	+ **EIDCo**	60.5	78.2	81.0	65.4	77.6	78.0	67.9	59.5	82.8	75.7	62.8	85.3	72.9
	+ FixMatch	62.0	79.2	81.2	65.8	78.0	78.4	68.1	60.2	83.2	76.7	65.0	86.3	73.7
	+ FixMatch + **EIDCo**	63.8	80.8	82.6	**71.5**	**80.1**	**80.9**	**72.1**	**61.3**	84.5	**78.6**	65.8	87.1	**75.8**
T	ViT-B	54.68	83.04	87.15	77.30	83.42	85.54	74.41	50.90	87.22	79.56	53.79	88.80	75.48
	CDTrans (ICLR'22) [82]	68.8	85.0	86.9	81.5	87.1	87.3	79.6	63.3	88.2	82.0	66.0	90.6	80.5
	TVT (WACV'23) [87]	74.89	86.82	89.47	82.78	87.95	88.27	79.81	71.94	90.13	85.46	74.62	90.56	83.56
	DOT-B (ACMMM'22) [47]	73.1	89.1	90.1	85.5	89.4	89.6	83.2	72.1	90.4	84.4	72.9	91.5	84.3
	SSRT (CVPR'22) [68]	75.17	88.98	91.09	85.13	88.29	89.95	85.04	74.23	91.26	85.70	**78.58**	91.78	85.43
	Baseline-B in SSRT	66.96	85.74	88.07	80.06	84.12	86.67	79.52	67.03	89.44	83.64	70.15	91.17	81.05
	+ **IEDCo**	69.42	86.04	88.59	81.71	84.87	87.15	80.90	68.89	90.08	84.55	71.21	91.34	82.06
	+ FixMatch	70.53	86.97	89.95	82.36	85.91	87.82	81.48	70.61	91.03	85.78	72.74	91.70	83.07
	+ FixMatch + **EIDCo**	**76.88**	**90.33**	**91.28**	**86.53**	**90.52**	**90.04**	**86.33**	**75.52**	**91.70**	**88.14**	77.06	**92.25**	**86.38**

tion (SSDA), and multi-source unsupervised domain adaptation (MSDA). For SSDA, a small number of labeled target samples (1-shot or 3-shot per class) are given [11]. In all three settings, our **EIDCo** loss considers unlabeled low-confidence target samples and source samples. Particularly, we regard multi-source domains as one in MSDA.

5.1. Datasets and Scenarios

Single Source UDA: *Office-Home* [73] is a challenging dataset with 15,500 images in 65 categories. It has four domains: Artistic, Clipart, Product, and Real-World (abbr. A, C, P, and R). *VisDA-2017* [54] is a large-scale dataset for synthetic-to-real adaptation. It contains 152,397 synthetic images and 55,388 real-world images in 12 categories.

Semi-Supervised DA: *DomainNet* [53] is initially a multi-source DA benchmark with 6 domains across 345 categories. For SSDA, four domains are commonly involved, *i.e.*, Real, Clipart, Painting, and Sketch (abbr. R, C, P, and S). Each of them contains images of 126 categories. *Office-Home* consists of A, C, P, and R domains with 65 classes.

Multi-Source UDA: *DomainNet* is a large-scale dataset that contains about 600,000 images in 345 categories, covering 6 domains with large domain gap: Clipart (C), Infograph (I), Painting (P), Quickdraw (Q), Real (R), and Sketch (S). We evaluate methods on five-source to one-target domain adaptation, resulting in 6 MSDA cases in total.

5.2. Implementation Details

For the baseline methods used in this paper, we choose GVB [9] and ToAlign [78] for SUDA, MME [58] and ToAlign [78] for SSDA, and ToAlign [78] for MSDA. To further evaluate the effectiveness with transformer backbone (*i.e.*, ViT [13]), we also adopt the baseline method in SSRT [68] (denoted by Baseline-B in SSRT) as baseline. For fair comparison, we report the average result of three random seeds.

For weak augmentation, we use random resize, crop, and horizontal flip. For strong augmentation, we adopt RandAugment [7] following FixMatch. For the training settings (*i.e*. batch size, optimizer, training iterations *etc*.), we directly follow the baseline method.

For the hyperparameters, we empirically set the probability threshold τ to 0.95 following FixMatch [65]. We find that $\tau = 0.95$ works well across all settings and tasks. For the temperature T_{co} in contrastive loss Eq. (3), we set it to 0.07 following previous contrastive learning methods [19, 86]. For the trade-off hyperparameters λ_{co} in Eq. (7), we directly set it to 1.0. For the size of the memory bank, we find that the proposed method is insensitive to it and adopt $|M| = 512$ for all experiments.

5.3. Comparison with State-of-the-Art Methods

Single Source UDA. Table 1 and Table 2 show the results on *Office-Home* and *VisDA-2017*, respectively. It

Table 2. Accuracy(%) of different UDA methods on *VisDA-2017* with ResNet-50 (R-50), ResNet-101 (R-101) and ViT-B (T). Following previous works, we report the instance-wise accuracy for ResNet-50, and the class-wise accuracy for ResNet-101 and ViT.

Net	Method	Acc	Method	Acc
R-50	Source-only	55.3	CDAN (NeurIPS'18) [44]	70.0
	DANN(ICML'15) [17]	57.4	SENTRY (ICCV'21) [55]	76.7
	MDD (ICML'19) [96]	74.6	CST (NeurIPS'21) [43]	80.6
	GVB (CVPR'20) [9]	75.3	ToAlign (NeurIPS'21) [78]	75.5
	+ **EIDCo**	77.3	+ **EIDCo**	78.0
	+ FixMatch	79.3	+ FixMatch	80.2
	+ FixMatch + **EIDCo**	82.0	+ FixMatch + **EIDCo**	**83.8**
R-101	Source-only	52.4	TSA (CVPR'21) [38]	78.6
	ADR (ICLR'18) [59]	74.8	MCC (ECCV'20) [23]	78.8
	SWD (CVPR'19) [31]	76.4	CRST+SUDA (CVPR'22) [94]	80.9
	BNM (CVPR'20) [8]	70.4	CRST+CaCo (CVPR'22) [21]	81.6
	CRST (ICCV'19) [102]	78.1	MCC+NWD (CVPR'22) [4]	83.7
	GVB (CVPR'20) [9]	77.5	ToAlign (NeurIPS'21) [78]	80.1
	+ **EIDCo**	79.7	+ **EIDCo**	82.9
	+ FixMatch	84.9	+ FixMatch	86.3
	+ FixMatch + **EIDCo**	87.0	+ FixMatch + **EIDCo**	**88.3**
T	ViT-B	72.6	TVT(WACV'23) [87]	83.92
	CDTrans (ICLR'22) [82]	88.4	SSRT (CVPR'22) [68]	88.76
	Baseline-B in SSRT [68]	85.23	+ FixMatch	86.61
	+ **EIDCo**	86.37	+ FixMatch + **EIDCo**	**89.84**

can be seen that our method can achieve great improvement based on different baseline methods and backbones. Based on the more powerful ViT model, the improvement over the strong baseline (*i.e.* baseline in SSRT + FixMatch) is still large (86.38% *vs.* 83.07%, 89.84% *vs.* 86.61%). The performance of baseline+**EIDCo** is lower than baseline+FixMatch. This is caused by the fact that high-confidence samples contain more discriminative and informative knowledge, and the proportion of these is gradually increased during the training process. There exists some inferiority of our method in certain adaptation cases compared with other SOTA methods. It is reasonable because different methods have advantages in certain adaptation scenarios and drawbacks in others, owing to the diversity of data distribution gaps. However, for average performance, our method always achieves SOTA.

Table 3. Accuracy(%) of different SSDA methods on *Office-Home* (*OH*) and *DomainNet* with ResNet34 (R-34) and ViT-B (T).

Net	Method	*OH*	*DomainNet*		Method	*OH*	*DomainNet*	
		3-shot	1-shot	3-shot		3-shot	1-shot	3-shot
R-34	S+Labeled-T	66.2	56.9	60.0	ATDOC (CVPR'21)	-	70.6	71.8
	HDAN (NeurIPS'20) [22]	-	69.5	71.3	CLDA (NeurIPS'21) [63]	75.5	71.9	75.3
	APE (ECCV'20) [28]	74.0	67.6	71.7	DECOTA (ICCV'21) [89]	-	-	75.6
	CDAC (CVPR'21) [33]	74.8	73.6	76.0	ECACL-P (ICCV'21) [35]	-	72.8	76.4
	STar (CVPR'21) [64]	-	70.0	73.2	MCL (IJCAI'22) [85]	77.1	74.4	76.5
	MME (ICCV'19) [58]	73.1	66.4	68.9	ToAlign (NeurIPS'21) [78]	75.1	70.6	73.0
	+ **EIDCo**	74.8	68.1	71.0	+ **EIDCo**	75.8	71.9	74.2
	+ FixMatch	75.5	72.3	74.8	+ FixMatch	76.4	72.7	75.4
	+ FixMatch + **EIDCo**	78.6	75.6	78.0	+ FixMatch + **EIDCo**	**79.0**	**75.8**	**78.3**
T	Baseline-B in SSRT [68]	85.3	77.6	80.2	+ FixMatch	88.0	84.8	85.7
	+ **EIDCo**	87.0	81.6	82.5	+ FixMatch + **EIDCo**	**89.3**	**86.2**	**87.1**

Semi-Supervised DA. Table 3 shows the mean accuracy on *Office-Home* and *DomainNet*, respectively. Combined with FixMatch, our method can achieve new SOTA performance on both tasks based on MME and ToAlign. It should be noted that CLDA [63] regards the classifier as a projection head, while we propose a novel Class Relationship Enhanced Features (CRF). It can be seen that our CRF consistently outperforms CLDA.

Table 4. Accuracy(%) of different MSDA methods on *DomainNet* with ResNet101 (R-101) and ViT-B (T).

Net	Method	Mean Acc	Method	Mean Acc
R-101	Source-Only	32.9	CMSS (ECCV'20) [88]	46.5
	ADDA (CVPR'17) [72]	32.2	HDAN (NeurIPS'20) [22]	47.6
	DCTN (CVPR'18) [81]	38.2	T-SVDNet (ICCV'21) [36]	47.0
	MCD (CVPR'18) [60]	38.5	LtC-MSDA (ECCV'20) [75]	47.4
	M^3SDA-β (ICCV'19)	42.6	PFSA (CVPR'21) [16]	48.5
	MLMSDA (Arxiv'20) [40]	44.3	STEM (ICCV'21) [50]	**53.4**
	ToAlign (NeurIPS'21) [78]	48.3	+ FixMatch	49.5
	+ **EIDCo**	48.9	+ FixMatch + **EIDCo**	51.7
T	Baseline-B in SSRT [68]	58.0	+ FixMatch	59.4
	+ **EIDCo**	58.8	+ FixMatch + **EIDCo**	**62.0**

Multi-Source UDA. Table 4 shows the result on *DomainNet*. For ResNet-101 backbone, our **EIDCo** loss brings about 2.2% improvement over the strong baseline. While other methods focus on selecting source samples or combining knowledge from multiple source domains, our method mainly explores the data structure in the target domain and exceeds most current methods. For ViT-B backbone, our method can still bring significant improvement.

5.4. Analysis and Discussion

Ablation studies of different components. Table 5 shows the ablation study results of different components in our proposed method. The experiment is conducted on UDA *Office-Home*. The first row (#0) shows the performance of GVB+FixMatch.

We first investigate the necessity of low-confidence samples in the original feature space. The results are shown in rows #1-3. It can be seen that using all samples (#1) only brings very limited improvements. Using high-confidence samples (#2) will slightly decrease the performance, indicating the category collision in *IDCo* loss needs to be carefully addressed. In contrast, using only low-confidence samples (#3) can boost the performance.

Rows #4-6 validate the effectiveness of our Class Relationship Enhanced Features (CRF). It can be seen that combining low-confidence samples with CRF can get the best result. The result in row #5 show that category collision still hinders feature learning of *IDCo* loss even with CRF.

Rows #7-8 validate the effect of target-dominated cross-domain mixup. Directly adding cross-domain mixup can bring 0.2% improvements. In this case, there are still high-confidence samples for *IDCo* loss. When adopting target-dominated mixup, only low-confidence samples are adopted in *IDCo* loss, and the performance can be further improved by 0.5%.

Table 5. Ablation studies of the proposed methods under the setting of UDA *Office-Home* with GVB baseline.

#	Samples for *IDCo* loss All	High	Low	CRF	Cross-Domain Mixup	Target Dominated	Acc
0							72.7
1	✓						73.0
2		✓					72.3
3			✓				73.3
4	✓			✓			73.4
5		✓		✓			72.4
6			✓	✓			74.4
7			✓	✓	✓		74.6
8			✓	✓	✓	✓	75.1

Ablation studies within CRF. Our proposed CRF can be seen as a boost version of original feature. Based on CRF, the classifier weights are involved in **EIDCo** loss, which can introduce semantic information. By adding class relationship embedded inter class term, the informative relationship between samples is further encoded. Table 6 show the superiority of our CRF compared with original features.

Table 6. Ablation studies of CRF in different datasets under UDA setting. For both datasets, we use ResNet-50 backbone.

#	Method	*Office-Home*	*VisDA-2017*
0	GVB+FixMatch	72.7	79.3
1	+ Our EIDCo w/ original feature	73.6	80.1
2	+ Our EIDCo w/ s_{cr}	75.1	82.0

Effectiveness on weaker baselines From the results on different domain adaptation tasks(*i.e.*, Table 1, 2, 3 and 4), we can see that FixMatch exceeds ours based on the same baseline, and the improvement is more significant when combined with FixMatch. This is because it uses high-confidence samples with more discriminative knowledge and less noise, and is complementary to ours. However, our method is effective on various baselines, and results on weaker ones are shown below. It can be seen that our EIDCo can bring consistent improvements on different baselines.

Table 7. Results on different baselines under UDA setting.

	Source only	+ EIDCo	DANN [17]	+ EIDCo	CDAN [44]	+ EIDCo
Office-Home	46.1	52.6(**+6.5**)	57.6	60.3 (**+2.7**)	65.8	67.6 (**+1.8**)
VisDA-2017	55.3	60.7 (**+5.4**)	57.4	61.9(**+4.5**)	70.0	73.8 (**+3.8**)

Visualization of features. To better understand the effect of CRF which can incorporate semantic information, we visualize the target domain features. As shown in Figure 4, the classifier weights in our CRF are closer to the features, indicating that the classifier weights are more representative of the learned features. We can also see that the target features of our method are more compact, which shows that our method can make *IDCo* loss effective for learning discriminative features.

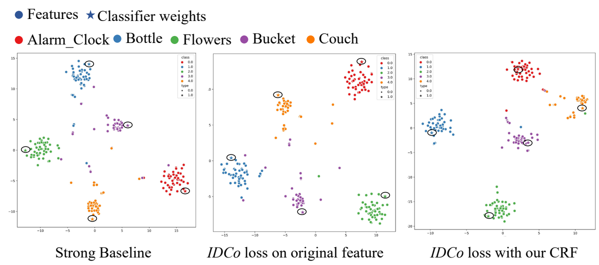

Figure 4. The t-SNE visualization of classifier weights and target domain features of different methods under the setting of UDA *Office-Home* C→A with GVB baseline. Best viewed in color.

Figure 5. "Alarm Clock" samples predicted by the strong baseline (*i.e.* GVB+FixMatch) and our method under the setting of UDA *Office-Home* C→R. Most predictions are the same, but our method can correct some hard samples (surrounded by blue boxes) compared with the strong baseline.

Visualization of samples. Here we provide visualization of the right predictions and failure cases of strong baseline (*i.e.* GVB+FixMatch) and our method. The results are shown in Figure 5. It can be seen that the failure cases have a different shape and appearance from source samples. The differences between strong baseline and our method are indicated by samples with blue boxes, which are misclassified by strong baseline but corrected by our method.

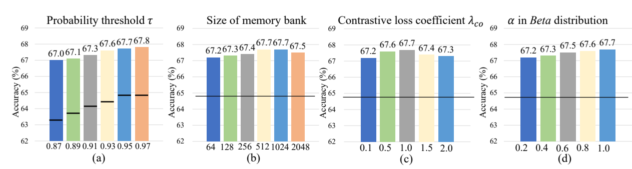

Figure 6. Sensitivity analysis of hyperparameters. The black lines represent accuracy of strong baseline (*i.e.* GVB+FixMatch).

Hyperparameter sensitivity. Here we analyze the hyperparameter sensitivity under the setting of UDA on *Office-Home* R→C with GVB baseline. We consider four hyperparameters, as shown in Figure 6. For the probability threshold τ in the baseline, our method can achieve similar accuracy around 0.95. The τ also affects the performance of strong baseline (*i.e.* GVB+FixMatch). It can be seen that our method is more stable than FixMatch given different τ. For other hyperparameters, it is evident that they are stable within specific ranges.

6. Conclusion

In this paper, we first analyze the limitations of instance discrimination contrastive loss for domain adaptation tasks, including the category collision and inadequate semantic information. For the first limitation, we propose to exploit low-confidence samples. For the second limitation, we propose to introduce domain-invariant and accurate semantic information through class relationship enhanced features and target-dominated cross-domain mixup. Extensive domain adaptation experiments show the effectiveness of proposed method, which achieves state-of-the-art performance.

7. Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant 62176246 and Grant 61836008. This work is also supported by Anhui Province Key Research and Development Plan (202304a05020045), Anhui Provincial Natural Science Foundation 2208085UD17, and the Fundamental Research Funds for the Central Universities (WK3490000006), and in part by STU Scientific Research Foundation for Talents (No.NTF22030).

References

[1] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine Learning*, 2010.

[2] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. In *NeurIPS*, 2007.

[3] Chaoqi Chen, Weiping Xie, Wenbing Huang, Yu Rong, Xinghao Ding, Yue Huang, Tingyang Xu, and Junzhou Huang. Progressive feature alignment for unsupervised domain adaptation. In *CVPR*, 2019.

[4] Lin Chen, Huaian Chen, Zhixiang Wei, Xin Jin, Xiao Tan, Yi Jin, and Enhong Chen. Reusing the task-specific classifier as a discriminator: Discriminator-free adversarial domain adaptation. In *CVPR*, 2022.

[5] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *TPAMI*, 2017.

[6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *ICML*, 2020.

[7] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *CVPR Workshops*, 2020.

[8] Shuhao Cui, Shuhui Wang, Junbao Zhuo, Liang Li, Qingming Huang, and Qi Tian. Towards discriminability and diversity: Batch nuclear-norm maximization under label insufficient situations. In *CVPR*, 2020.

[9] Shuhao Cui, Shuhui Wang, Junbao Zhuo, Chi Su, Qingming Huang, and Qi Tian. Gradually vanishing bridge for adversarial domain adaptation. In *CVPR*, 2020.

[10] Zhijie Deng, Yucen Luo, and Jun Zhu. Cluster alignment with a teacher for unsupervised domain adaptation. In *ICCV*, 2019.

[11] Jeff Donahue, Judy Hoffman, Erik Rodner, Kate Saenko, and Trevor Darrell. Semi-supervised domain adaptation with instance constraints. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013.

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

[13] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2020.

[14] Debidatta Dwibedi, Yusuf Aytar, Jonathan Tompson, Pierre Sermanet, and Andrew Zisserman. With a Little Help From My Friends: Nearest-Neighbor Contrastive Learning of Visual Representations. In *ICCV*, 2021.

[15] Geoffrey French, Michal Mackiewicz, and Mark Fisher. Self-ensembling for visual domain adaptation. In *ICLR*, 2018.

[16] Yangye Fu, Ming Zhang, Xing Xu, Zuo Cao, Chao Ma, Yanli Ji, Kai Zuo, and Huimin Lu. Partial feature selection and alignment for multi-source domain adaptation. In *CVPR*, 2021.

[17] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *ICML*, 2015.

[18] Zhiqiang Gao, Shufei Zhang, Kaizhu Huang, Qiufeng Wang, and Chaoliang Zhong. Gradient distribution alignment certificates better adversarial domain adaptation. In *ICCV*, 2021.

[19] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, 2020.

[20] Dapeng Hu, Jian Liang, Qibin Hou, Hanshu Yan, and Yunpeng Chen. Adversarial domain adaptation with prototype-based normalized output conditioner. *IEEE Transactions on Image Processing*, 2021.

[21] Jiaxing Huang, Dayan Guan, Aoran Xiao, Shijian Lu, and Ling Shao. Category contrast for unsupervised domain adaptation in visual tasks. In *CVPR*, 2022.

[22] Xuan Jin, Shuhui Wang, Yuan He, Qingming Huang, et al. Heuristic domain adaptation. In *NeurIPS*, 2020.

[23] Ying Jin, Ximei Wang, Mingsheng Long, and Jianmin Wang. Minimum class confusion for versatile domain adaptation. In *ECCV*, 2020.

[24] Yannis Kalantidis, Mert Bulent Sariyildiz, Noe Pion, Philippe Weinzaepfel, and Diane Larlus. Hard negative mixing for contrastive learning. In *NeurIPS*, 2020.

[25] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. In *CVPR*, 2019.

[26] Donghyun Kim, Kuniaki Saito, Tae-Hyun Oh, Bryan A Plummer, Stan Sclaroff, and Kate Saenko. Cross-domain self-supervised learning for domain adaptation with few source labels. *arXiv preprint arXiv:2003.08264*, 2020.

[27] Seunghyeon Kim, Jaehoon Choi, Taekyung Kim, and Changick Kim. Self-training and adversarial background regularization for unsupervised domain adaptive one-stage object detection. In *CVPR*, 2019.

[28] Taekyung Kim and Changick Kim. Attract, perturb, and explore: Learning a feature alignment network for semi-supervised domain adaptation. In *ECCV*, 2020.

[29] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NeurIPS*, 2012.

[30] Ananya Kumar, Tengyu Ma, and Percy Liang. Understanding self-training for gradual domain adaptation. In *ICML*, 2020.

[31] Chen-Yu Lee, Tanmay Batra, Mohammad Haris Baig, and Daniel Ulbricht. Sliced wasserstein discrepancy for unsupervised domain adaptation. In *CVPR*, 2019.

[32] Kibok Lee, Yian Zhu, Kihyuk Sohn, Chun-Liang Li, Jinwoo Shin, and Honglak Lee. i-mix: A domain-agnostic strategy for contrastive representation learning. In *ICLR*, 2021.

[33] Jichang Li, Guanbin Li, Yemin Shi, and Yizhou Yu. Cross-domain adaptive clustering for semi-supervised domain adaptation. In *CVPR*, 2021.

[34] Junjie Li, Yixin Zhang, Zilei Wang, and Keyu Tu. Probabilistic contrastive learning for domain adaptation. *arXiv preprint arXiv:2111.06021*, 2021.

[35] Kai Li, Chang Liu, Handong Zhao, Yulun Zhang, and Yun Fu. Ecacl: A holistic framework for semi-supervised domain adaptation. In *ICCV*, 2021.

[36] Ruihuang Li, Xu Jia, Jianzhong He, Shuaijun Chen, and Qinghua Hu. T-svdnet: Exploring high-order prototypical correlations for multi-source domain adaptation. In *ICCV*, 2021.

[37] Shuang Li, Fangrui Lv, Binhui Xie, Chi Harold Liu, Jian Liang, and Chen Qin. Bi-classifier determinacy maximization for unsupervised domain adaptation. In *AAAI*, 2021.

[38] Shuang Li, Mixue Xie, Kaixiong Gong, Chi Harold Liu, Yulin Wang, and Wei Li. Transferable semantic augmentation for domain adaptation. In *CVPR*, 2021.

[39] Shuang Li, Mixue Xie, Fangrui Lv, Chi Harold Liu, Jian Liang, Chen Qin, and Wei Li. Semantic concentration for domain adaptation. In *ICCV*, 2021.

[40] Zhenpeng Li, Zhen Zhao, Yuhong Guo, Haifeng Shen, and Jieping Ye. Mutual learning network for multi-source domain adaptation. *arXiv preprint arXiv:2003.12944*, 2020.

[41] Jian Liang, Dapeng Hu, and Jiashi Feng. Domain adaptation with auxiliary target domain-oriented classifier. In *CVPR*, 2021.

[42] Hong Liu, Mingsheng Long, Jianmin Wang, and Michael Jordan. Transferable adversarial training: A general approach to adapting deep classifiers. In *ICML*, 2019.

[43] Hong Liu, Jianmin Wang, and Mingsheng Long. Cycle self-training for domain adaptation. In *NeurIPS*, 2021.

[44] Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. In *NeurIPS*, 2018.

[45] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. In *ICML*, 2017.

[46] You-Wei Luo and Chuan-Xian Ren. Conditional bures metric for domain adaptation. In *CVPR*, 2021.

[47] Wenxuan Ma, Jinming Zhang, Shuang Li, Chi Harold Liu, Yulin Wang, and Wei Li. Making the best of both worlds: A domain-oriented transformer for unsupervised domain adaptation. In *ACMMM*, 2022.

[48] Ke Mei, Chuang Zhu, Jiaqi Zou, and Shanghang Zhang. Instance adaptive self-training for unsupervised domain adaptation. In *ECCV*, 2020.

[49] Jaemin Na, Heechul Jung, Hyung Jin Chang, and Wonjun Hwang. Fixbi: Bridging domain spaces for unsupervised domain adaptation. In *CVPR*, 2021.

[50] Van-Anh Nguyen, Tuan Nguyen, Trung Le, Quan Hung Tran, and Dinh Phung. Stem: An approach to multi-source domain adaptation with guarantees. In *ICCV*, 2021.

[51] Fei Pan, Inkyu Shin, Francois Rameau, Seokju Lee, and In So Kweon. Unsupervised intra-domain adaptation for semantic segmentation through self-supervision. In *CVPR*, 2020.

[52] Yingwei Pan, Ting Yao, Yehao Li, Yu Wang, Chong-Wah Ngo, and Tao Mei. Transferrable prototypical networks for unsupervised domain adaptation. In *CVPR*, 2019.

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

[54] Xingchao Peng, Ben Usman, Neela Kaushik, Judy Hoffman, Dequan Wang, and Kate Saenko. Visda: The visual domain adaptation challenge. In *arXiv preprint arXiv:1710.06924*, 2017.

[55] Viraj Prabhu, Shivam Khare, Deeksha Kartik, and Judy Hoffman. Sentry: Selective entropy optimization via committee consistency for unsupervised domain adaptation. In *ICCV*, 2021.

[56] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *NeurIPS*, 2015.

[57] Aadarsh Sahoo, Rutav Shah, Rameswar Panda, Kate Saenko, and Abir Das. Contrast and mix: Temporal contrastive video domain adaptation with background mixing. In *NeurIPS*, 2021.

[58] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, Trevor Darrell, and Kate Saenko. Semi-supervised domain adaptation via minimax entropy. In *ICCV*, 2019.

[59] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Adversarial dropout regularization. In *ICLR*, 2018.

[60] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *CVPR*, 2018.

[61] Nikunj Saunshi, Orestis Plevrakis, Sanjeev Arora, Mikhail Khodak, and Hrishikesh Khandeparkar. A theoretical analysis of contrastive unsupervised representation learning. In *ICML*. PMLR, 2019.

[62] Zhiqiang Shen, Zechun Liu, Zhuang Liu, Marios Savvides, Trevor Darrell, and Eric Xing. Un-mix: Rethinking image mixtures for unsupervised visual representation learning. *arXiv preprint arXiv:2003.05438*, 2020.

[63] Ankit Singh. Clda: Contrastive learning for semi-supervised domain adaptation. In *NeurIPS*, 2021.

[64] Anurag Singh, Naren Doraiswamy, Sawa Takamuku, Megh Bhalerao, Titir Dutta, Soma Biswas, Aditya Chepuri, Balasubramanian Vengatesan, and Naotake Natori. Improving semi-supervised domain adaptation using effective target selection and semantics. In *CVPR*, 2021.

[65] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. In *NeurIPS*, 2020.

[66] Baochen Sun, Jiashi Feng, and Kate Saenko. Return of frustratingly easy domain adaptation. In *AAAI*, 2016.

[67] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *ECCV*, 2016.

[68] Tao Sun, Cheng Lu, Tianshuo Zhang, and Haibin Ling. Safe self-refinement for transformer-based domain adaptation. In *CVPR*, 2022.

[69] Hui Tang, Ke Chen, and Kui Jia. Unsupervised domain adaptation via structurally regularized deep clustering. In *CVPR*, 2020.

[70] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *NeurIPS*, 2017.

[71] Michael Tschannen, Josip Djolonga, Paul K Rubenstein, Sylvain Gelly, and Mario Lucic. On mutual information maximization for representation learning. *arXiv preprint arXiv:1907.13625*, 2019.

[72] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *CVPR*, 2017.

[73] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, , and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *CVPR*, 2017.

[74] Feng Wang and Huaping Liu. Understanding the behaviour of contrastive loss. In *CVPR*, 2021.

[75] Hang Wang, Minghao Xu, Bingbing Ni, and Wenjun Zhang. Learning to combine: Knowledge aggregation for multi-source domain adaptation. In *ECCV*, 2020.

[76] Ximei Wang, Liang Li, Weirui Ye, Mingsheng Long, and Jianmin Wang. Transferable attention for domain adaptation. In *AAAI*, 2019.

[77] Guoqiang Wei, Cuiling Lan, Wenjun Zeng, and Zhibo Chen. Metaalign: Coordinating domain alignment and classification for unsupervised domain adaptation. In *CVPR*, 2021.

[78] Guoqiang Wei, Cuiling Lan, Wenjun Zeng, Zhizheng Zhang, and Zhibo Chen. Toalign: Task-oriented alignment for unsupervised domain adaptation. In *NeurIPS*, 2021.

[79] Yuan Wu, Diana Inkpen, and Ahmed El-Roby. Dual mixup regularized learning for adversarial domain adaptation. In *ECCV*, 2020.

[80] Minghao Xu, Jian Zhang, Bingbing Ni, Teng Li, Chengjie Wang, Qi Tian, and Wenjun Zhang. Adversarial domain adaptation with domain mixup. In *AAAI*, 2020.

[81] Ruijia Xu, Ziliang Chen, Wangmeng Zuo, Junjie Yan, and Liang Lin. Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2018.

[82] Tongkun Xu, Weihua Chen, WANG Pichao, Fan Wang, Hao Li, and Rong Jin. Cdtrans: Cross-domain transformer for unsupervised domain adaptation. In *International Conference on Learning Representations*, 2021.

[83] Hongliang Yan, Yukang Ding, Peihua Li, Qilong Wang, Yong Xu, and Wangmeng Zuo. Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation. In *CVPR*, 2017.

[84] Shen Yan, Huan Song, Nanxiang Li, Lincan Zou, and Liu Ren. Improve unsupervised domain adaptation with mixup training. *arXiv preprint arXiv:2001.00677*, 2020.

[85] Zizheng Yan, Yushuang Wu, Guanbin Li, Yipeng Qin, Xiaoguang Han, and Shuguang Cui. Multi-level consistency learning for semi-supervised domain adaptation. In *IJCAI*, 2022.

[86] Fan Yang, Kai Wu, Shuyi Zhang, Guannan Jiang, Yong Liu, Feng Zheng, Wei Zhang, Chengjie Wang, and Long Zeng. Class-aware contrastive semi-supervised learning. In *CVPR*, 2022.

[87] Jinyu Yang, Jingjing Liu, Ning Xu, and Junzhou Huang. Tvt: Transferable vision transformer for unsupervised domain adaptation. In *WACV*, 2023.

[88] Luyu Yang, Yogesh Balaji, Ser-Nam Lim, and Abhinav Shrivastava. Curriculum manager for source selection in multi-source domain adaptation. In *Eur. Conf. Comput. Vis.*, 2020.

[89] Luyu Yang, Yan Wang, Mingfei Gao, Abhinav Shrivastava, Kilian Q Weinberger, Wei-Lun Chao, and Ser-Nam Lim. Deep co-training with task decomposition for semi-supervised domain adaptation. In *ICCV*, 2021.

[90] Shiqi Yang, Yaxing Wang, Kai Wang, Shangling Jui, and Joost van de Weijer. Attracting and dispersing: A simple approach for source-free domain adaptation. In *NeurIPS*, 2022.

[91] Xiangyu Yue, Zangwei Zheng, Shanghang Zhang, Yang Gao, Trevor Darrell, Kurt Keutzer, and Alberto Sangiovanni Vincentelli. Prototypical cross-domain self-supervised learning for few-shot unsupervised domain adaptation. In *CVPR*, 2021.

[92] Zhongqi Yue, Qianru Sun, Xian-Sheng Hua, and Hanwang Zhang. Transporting causal mechanisms for unsupervised domain adaptation. In *ICCV*, 2021.

[93] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. In *ICLR*, 2018.

[94] Jingyi Zhang, Jiaxing Huang, Zichen Tian, and Shijian Lu. Spectral unsupervised domain adaptation for visual recognition. In *CVPR*, 2022.

[95] Pan Zhang, Bo Zhang, Ting Zhang, Dong Chen, Yong Wang, and Fang Wen. Prototypical pseudo label denoising

and target structure learning for domain adaptive semantic segmentation. In *CVPR*, 2021.

[96] Yuchen Zhang, Tianle Liu, Mingsheng Long, and Michael Jordan. Bridging theory and algorithm for domain adaptation. In *ICML*, 2019.

[97] Yixin Zhang and Zilei Wang. Joint adversarial learning for domain adaptation in semantic segmentation. In *AAAI*, 2020.

[98] Yixin Zhang, Zilei Wang, and Yushi Mao. Rpn prototype alignment for domain adaptive object detector. In *CVPR*, 2021.

[99] Rui Zhu, Bingchen Zhao, Jingen Liu, Zhenglong Sun, and Chang Wen Chen. Improving contrastive learning by visualizing feature transformation. In *ICCV*, 2021.

[100] Xinge Zhu, Hui Zhou, Ceyuan Yang, Jianping Shi, and Dahua Lin. Penalizing top performers: Conservative loss for semantic segmentation adaptation. In *ECCV*, 2018.

[101] Yang Zou, Zhiding Yu, BVK Kumar, and Jinsong Wang. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training. In *ECCV*, 2018.

[102] Yang Zou, Zhiding Yu, Xiaofeng Liu, BVK Kumar, and Jinsong Wang. Confidence regularized self-training. In *ICCV*, 2019.