UFDA: Universal Federated Domain Adaptation with Practical Assumptions

Xinhui Liu[1,2], Zhenghao Chen[2], Luping Zhou[2], Dong Xu[3], Wei Xi[1*],
Gairui Bai[1], Yihan Zhao[1], Jizhong Zhao[1]

[1]School of Computer Science and Technology, Xi'an Jiaotong University, Xi'an, China
[2]School of Electrical and Computer Engineering, The University of Sydney, Sydney, Australia
[3]Department of Computer Science, The University of Hong Kong, Hong Kong SAR, China
liuxinhui@stu.xjtu.edu.cn, zhenghao.chen@sydney.edu.au, xiwei@xjtu.edu.cn

Abstract

Conventional Federated Domain Adaptation (FDA) approaches usually demand an abundance of assumptions, which makes them significantly less feasible for real-world situations and introduces security hazards. This paper relaxes the assumptions from previous FDAs and studies a more practical scenario named Universal Federated Domain Adaptation (UFDA). It only requires the black-box model and the label set information of each source domain, while the label sets of different source domains could be inconsistent, and the target-domain label set is totally blind. Towards a more effective solution for our newly proposed UFDA scenario, we propose a corresponding methodology called Hot-Learning with Contrastive Label Disambiguation (HCLD). It particularly tackles UFDA's domain shifts and category gaps problems by using one-hot outputs from the black-box models of various source domains. Moreover, to better distinguish the shared and unknown classes, we further present a cluster-level strategy named Mutual-Voting Decision (MVD) to extract robust consensus knowledge across peer classes from both source and target domains. Extensive experiments on three benchmark datasets demonstrate that our method achieves comparable performance for our UFDA scenario with much fewer assumptions, compared to previous methodologies with comprehensive additional assumptions.

Introduction

Federated Learning (FL) (McMahan et al. 2017; Mohassel and Zhang 2017; Mohassel and Rindal 2018) allows models to be optimized across decentralized devices while keeping data localized, where no clients are required to share their local confidential data with other clients or the centralized server. Traditional FL often struggles to produce models that can effectively generalize to new unlabeled domains from clients due to the barrier presented by domain shifts (Yang et al. 2019). To address this, Federated Domain Adaptation (FDA) (Fantauzzo et al. 2022; Gilad-Bachrach et al. 2016) are proposed and achieved tremendous success as it allows knowledge transfer from decentralized source domains to an unlabeled target domain using Domain Adaption (DA) techniques. Nonetheless, current FDA scenarios often operate under the presumption that model parameters or gra-
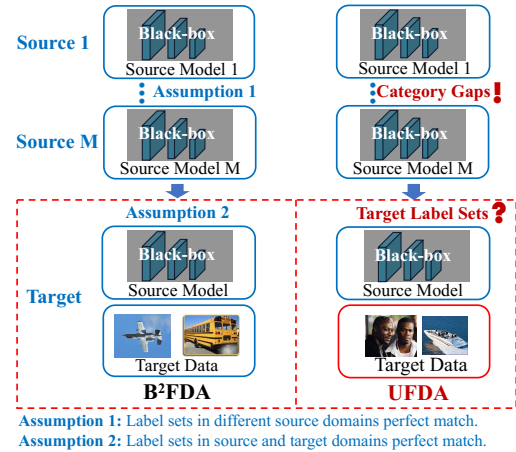

Figure 1: Overview of Federated Domain Adaptation for Black-Box Models (B²FDA) (left), and our proposed Universal Federated Domain Adaptation (UFDA) (right). Different than B²FDA, where the label set consistency among different source domains (*i.e.*, Assumption 1) and between source and target domains (*i.e.*, Assumption 2) are required, our UFDA scenario allows the label set diversity of source domains and the target domain.

dients are optimized based on the source domain. However, acquiring such information in real-world situations is exceptionally challenging due to commercial confidentiality. Also, exposing such information introduces potential risks such as model misuse and white-box attacks.

To establish a relaxed condition, Federated Domain Adaptation with Black-Box Models (B²FDA) (Wu et al. 2021; Liang et al. 2022; Liu et al. 2023) is introduced, where the target-domain client can only access the application programming interfaces (APIs) of various source domains. However, most existing B²FDA approaches assume that the label sets of different source domains must perfectly align with each other and that of the target domains. This assumption is particularly challenging to fulfill in real-world scenarios. First, source data can originate from vastly diverse domains. For example, the biometric data of a single client could stem from unrelated sources like the medical domain (*e.g.*, clinical records from different hospitals)

or the financial domain (*e.g.*, user records from different Banks). Second, in real-world scenarios, acquiring information about the label set of the target domain samples is often a formidable task. Consequently, attempting to align the label sets of source and target domains becomes impractical.

To further minimize those in-practical assumptions from B^2FDA, we introduce a new scenario Universal Federated Domain Adaptation (UFDA) towards a practical FDA setup with practical assumptions. As shown in Figure. 1, in UFDA, the target domain solely requires a black-box model, devoid of its specifics (*e.g.*, gradients). Meanwhile, we only need to know the source domains' label sets, which are not required to be identical as in B^2FDA scenarios, and the target domain's label set will remain entirely unknown as most real-world DA scenarios. On the other hand, different from most existing B^2FDA setups (Liang et al. 2022), our UFDA presents two unique challenges: First, as the target domain's label set is completely unknown, the model optimized based on each individual source domain could be particularly imprecise for those unique categories of the target domain. Second, the completion uncertainty of the target domain's label set also makes it impossible to distinguish the shared and unknown classes among source and target domains. However, it is important in FDA problems to guarantee the consistency of label sets between source and target domains.

To tackle the first challenge, we propose a methodology called Hot-Learning with Contrastive Label Disambiguation (HCLD). It adopts one-hot outputs (without confidence) produced by various source APIs, which generate more than one candidate pseudo-labels for each target sample. Compared with previous FDA methods, which directly adopt one candidate (with confidence) from source APIs by using the probability function (*e.g.*, Softmax), our method can mitigate the impact caused by the falsely higher confidence in these non-existent categories. To obtain more credible pseudo-labels, we propose a Gaussian Mixture Model (GMM) based Contrastive Label Disambiguation (GCLD) method, which sharpens the shared-class confidence and smooths the unknown-class confidence. Specifically, it leverages Contrastive Learning (CL) (Khosla et al. 2020) strategy to dynamically generate prototype-based clustering, which will fit a GMM (Permuter, Francos, and Jermyn 2006) based on its self-entropy distribution for sample divisions. Therefore, the easy-to-learn sample can be treated as a shared-class sample while the hard-to-learn sample can be treated as an unknown-class sample. Furthermore, to address the second challenge, we propose a cluster-level Mutual-Voting Decision (MVD) strategy by leveraging the consensus knowledge of shared classes among source and target domains. We calculate a "mutual voting score" for each class based on the overlapping samples recognized as the same category from all APIs (*i.e.*, source + target). Then, we use this score to distinguish each class as "shared" or "unknown" type.

Our contributions are summarized as follows:

- We introduce a new FDA scenario, UFDA, which not only inherits relaxed assumptions as in B^2FDA, but also eliminates the consistency requirement of label sets

among source domains and keeps the target domain's label sets completely unknown, towards a practical scenario for real-world situations.

- We proposed a novel methodology, HCLD, to address the imprecision issue for samples from non-existent categories. It adopts ensemble one-hot outputs from multi-source APIs to produce multiple candidate pseudo-labels and uses a GMM-based strategy GCLD to disambiguate those candidates.

- We present a cluster-level strategy MVD to distinguish shared and unknown classes by leveraging consensus knowledge across peer classes from source and target domains.

- We conduct extensive experiments on three DA benchmarks. The results demonstrate that our method exhibits performance on par with previous MDA approaches, yet relies on significantly fewer assumptions. This substantiates the practicality of our method.

Related Works

Multi-source Domain Adaptation (MDA)

MDA (Ben-David et al. 2010; Blitzer et al. 2007; Liu et al. 2022) has gained significant attention as a means to mitigate performance degradation caused by domain shifts. Despite the achievements of MDA, many existing approaches (Hoffman, Mohri, and Zhang 2018; Zhao et al. 2018; Chen et al. 2017, 2022b) are limited to the assumption of perfectly matched label sets and have to access the raw multi-source, which can be inefficient and may raise concerns regarding data protection policies (Voigt and Von dem Bussche 2017).

To tackle category shift issues, the UniMDA scenario is introduced (Xu et al. 2018; Kundu et al. 2020; Shui et al. 2021, 2022; Saito and Saenko 2021; Shan, Ma, and Wen 2023; Chen et al. 2022a), where the label set among multi-sources differ, and no prior knowledge about the target label sets is accessible. In UniMDA, the concept of category shift was first introduced in DCTN (Xu et al. 2018), which acknowledged that the number of categories in each source domain may differ from the target domain. DCTN learns transferable and discriminative representations via an alternating adaptation algorithm and a distribution-weighted combining rule. To address data privacy issues, source-free domain adaptation (SFDA) (Liang, Hu, and Feng 2020; Ahmed et al. 2021; Dong et al. 2021; Tian et al. 2022; Zhao et al. 2022) and federated domain adaptation (FDA) (Li et al. 2020; Peng et al. 2019; Feng et al. 2021; Wu and Gong 2021) have attracted increasing attention. Instead of accessing the raw data directly, SFDA utilizing the well-trained model rather than the raw labeled data has emerged as a possible solution to this problem. Another setting that deals with unavailable source data is the FDA, where the goal is to develop a global model from decentralized datasets by aggregating the parameters of each local client (Csurka et al. 2022). Inspired by FL, (Peng et al. 2019) first raised the concept of the FDA. This work provides a solution named Federated Adversarial Domain Adaptation, which aims to address the FDA problem in a federated learning system using adversarial techniques.
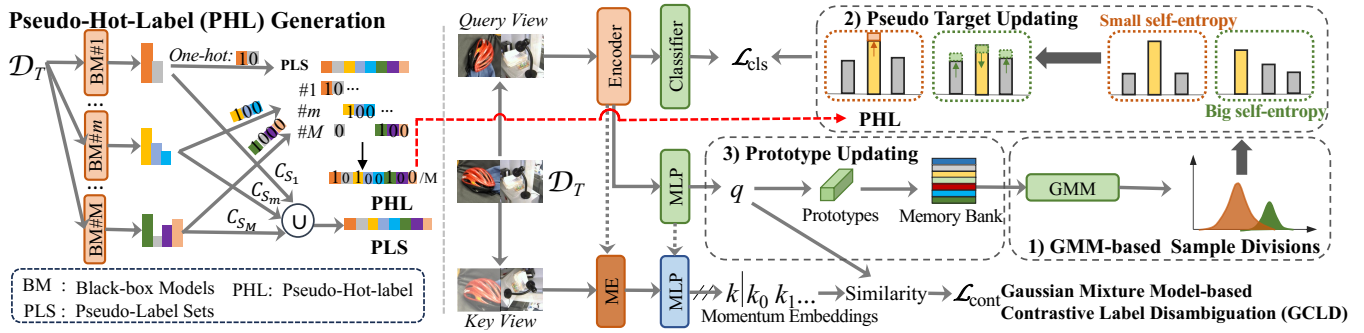
Figure 2: Our Hot-Learning with Contrastive Label Disambiguation (HCLD) methodology contains two key components: Pseudo-Hot-Label (PHL) Generation and Gaussian Mixture Model-based Contrastive Label Disambiguation (GCLD). $\mathcal{D}_T$, ME, '//', GMM, and MLP respectively represent the unlabeled target dataset, Momentum Embeddings maintained by a queue structure, the stop-gradient operation, Gaussian Mixture Model, and the Multi-Layered Perceptron module.

However, these approaches do not address both of these limitations simultaneously. Recently, a few works (Kundu et al. 2020; Saito et al. 2020; Qu et al. 2023) explore SFDA under category shift. Despite their effectiveness, they require dedicated multi-source model specifics, which can be restricted due to their commercial value and associated risks, such as model misuse and white-box attacks. In this work, we deal with a practical scenario of UFDA, which requires neither the shared data and model specifics, consistency of label sets among source domains, and information on the target domain label set.

Contrastive Learning (CL)

Due to the success of CL (Wu et al. 2018; Chen et al. 2020, 2021, 2022c; Guo et al. 2023), numerous efforts have been made to improve the robustness of classification tasks by harnessing the advantages of CL. For instance, (Zheltonozhskii et al. 2022) employed CL as a pre-training technique for their classification model. Another approach, RRL (Li, Xiong, and Hoi 2021) introduced label cleaning utilizing two thresholds on soft labels, which are calculated from the predictions of previous epochs and their nearest neighbors. Similarly, Sel-CL (Li et al. 2022) leveraged nearest neighbors to select confident pairs for supervised CL (Khosla et al. 2020). Despite their demonstrated effectiveness, these methods are not explicitly designed to tackle the category shift between the noise-label sets and the ground-truth label set.

Methodology

Preliminaries

We are given M source datasets from different clients $\{\mathcal{D}_S^m\}_{m=1}^M$ and an unlabeled target client $\mathcal{D}_T$, where each source client contains N_m labeled source samples $\mathcal{D}_S^m := \{(x_i^m, y_i^m)\}_{i=1}^{N_m}$ and the target client comprising N_T unlabeled samples $\{x_i\}_{i=1}^{N_T}$, $s.t., x_i \in X^T$. In most real-world scenarios, each client's data and model specifics are stored exclusively on local systems, ensuring that they are not shared with other clients or a centralized server. Therefore, the label sets between the aforementioned multi-source and target clients may exhibit significant variations.

While, most existing FDA studies intuitively assume that multi-source and target clients share the same label sets, which is not practical. Inspired by the research of UniMDA, we define $\mathcal{C}_{s_m}$ as the label sets for the m-th source node and $\mathcal{C}_t$ as the label set for the target node. The label sets $\mathcal{C}_m$ represents the common labels between $\mathcal{C}_{s_m}$ and $\mathcal{C}_t$. Furthermore, $\overline{\mathcal{C}}_{s_m} = \mathcal{C}_{s_m} \backslash \mathcal{C}_m$ represent the label sets exclusive to $\mathcal{D}_S^m$. Similarly, $\overline{\mathcal{C}}_t = \mathcal{C}_t \backslash \{\cup_m \mathcal{C}_m\}$ indicates the classes in the target domain $\mathcal{D}_T$ that are unknown in the multi-source domains, as they should never appear in any source label sets. The label sets $\mathcal{C}$ represent the union of shared classes, $i.e.$, $\mathcal{C} = \cup_m \mathcal{C}_m$. It is important to note that the target data are fully unlabeled and the target label set (which is inaccessible during training) is only used to define the UFDA problem.

HCLD

Our proposed HCLD aims to establish an effective mapping that can accurately classify target samples if they correspond to the shared class $\mathcal{C}$, or confuse the samples with an "unknown" class. As shown in Figure. 2, HCLD consists of two key components: 1) Pseudo-Hot-Label (PHL) Generation; 2) Gaussian Mixture Model-based Contrastive Label Disambiguation (GCLD). Firstly, to mitigate the impact caused by multi-source APIs' falsely higher confidence for the non-existent categories, we calculate the pseudo-labels for each target sample with the proposed PHL Generation strategy. Then, we adopt the GCLD manner to obtain more credible pseudo-labels, which sharpens the shared-class confidence and smooths the unknown-class confidence.

PHL Generation In UFDA, only the label sets $\{\mathcal{C}_{s_m}\}_{m=1}^M$ in each source domain and the softmax output Y_m^T in each source APIs for target samples are acceptable for the target party: $Y_m^T = f_S^m(X^T)$. Considering the domain shift between multiple source and target domains, the key challenge lies in obtaining more reliable pseudo-labels for each target sample. Empirically, individual source APIs often display increased confidence levels for both shared- and non-existent categories. Such a trend adversely affects the accuracy of pseudo-labels that are produced using these confi-

dence scores.

To address the aforementioned limitation, we suggest the use of an ensemble of multiple one-hot outputs to create the pseudo-labels, referred to as PHL C_{pse} (*i.e.,* Figure. 2), which generates multiple candidate pseudo-labels for each target sample, providing a broader and potentially more accurate range of labeling options. Given the lack of pre-existing knowledge about the target label sets, we determine the Pseudo-Label Sets (PLS) for the target domain by the following method: $\hat{\mathcal{C}}_T = \cup_m \mathcal{C}_{s_m}$. This strategy ensures that the accurate labels identified by each APIs are encompassed within these candidate pseudo-labels.

GCLD The candidate pseudo-labels in the above PHL C_{pse} inevitably contain unknown categories due to the gap between multi-source and target domains. We adopt GCLD, which iteratively sharpens the possible shared-class confidence, smooths the possible unknown-class confidence, and obtains more credible pseudo-labels.

The critical challenge is distinguishing between shared- and unknown-class samples. Inspired by (Permuter, Francos, and Jermyn 2006), GMM can better distinguish clean and noisy samples due to its flexibility in the sharpness of distribution. Treating easy-to-learn samples as shared class instances and challenging samples as unknown-class instances, we facilitate the acquisition of discriminative image representations through CL and construct a GMM over the representations for sample divisions. Typically, the dimension of contrastive prototypes is limited by the pseudo-label sets, making it difficult for GMM to handle this scenario effectively. Therefore, we utilize a comprehensive Memory Bank denoted $U^e = \{u_1^e, \ldots, u_{N_T}^e\}$ that maintains the running average of the features of all target samples. Here, U^e represents the Memory Bank in epoch e. We initialize U^e with random unit vectors and update its values by mixing U^e and U^{e-1} during training (details in the next subsection).

$$U^e \leftarrow \delta U^e + (1 - \delta) U^{e-1} \tag{1}$$

where δ is a mixing parameter. The self-entropy of U^e for each sample can be defined as:

$$l_{ce}(i) = -\sum u_i^e \log (u_i^e), i \in \{1, \ldots, N_T\} \tag{2}$$

1) GMM-based Sample Divisions. To distinguish between shared- and unknown-class samples, we fit a two-component GMM to the self-entropy distribution l_{ce} using the Expectation Maximization algorithm. Each sample is assigned a shared probability w_i, which is the posterior probability $p(\theta \mid \ell_{ce})$, where θ corresponds to the Gaussian component with a smaller mean (indicating a smaller self-entropy). Based on the shared probability, we divide all target samples into two sets: W^1 (the sample may with shared class) and $W^0 = D_T \backslash W^1$ (the sample may with unknown class) by setting a threshold σ.

2) Pseudo Target Updating. In terms of the above distinguished shared class W^1 and unknown class samples W^0, we sharpen the shared-class confidence and smooth the unknown-class confidence to update the pseudo-labels C_{pse} as follows,

$$C_{pse}^e \leftarrow \phi(\phi C_{pse}^e + (1 - \phi) C_{pse}^{e-1}) + (1 - \phi) z^e \tag{3}$$

$$z^e = \begin{cases} Onehot(u_i^e) & x_i \in W^1 \\ 1/n_C & \text{otherwise.} \end{cases} \tag{4}$$

where ϕ is a tunable hyperparameter. Ultimately, the pseudo-labels with shared classes will be clustered around each cluster center, while confusing the pseudo-labels with unknown labels.

3) Prototype Updating. Since the contrastive loss induces a clustering effect in the embedding space, we maintain a prototype embedding vector μ_c corresponding to each class in $\hat{\mathcal{C}}_T$, which serves as a set of representative embedding vectors. This approach adopts updating μ_c with a moving-average style:

$$\begin{aligned} \mu_c &= \text{Normalize} \left(\gamma \mu_c + (1 - \gamma) q \right), \\ & \text{if } c = \arg\max_{j \in \hat{\mathcal{C}}_T} f^j \left(\text{Aug}_q(x) \right) \end{aligned} \tag{5}$$

where the momentum parameter γ was set as 0.99. Then, we iteratively update the above-mentioned Memory Bank U^e with the moving-updating mechanism $U^e \leftarrow q * \mu_c^T$ (details of q in the next subsection).

Training Objective Given the target samples with PHL $\{x^i, c_{pse}^i\}_{i=1}^{N_T}$, we generate a query view $Aug_q(x)$ and a key view $Aug_k(x)$ with the randomized data augmentation $Aug(x)$. Then, HCLD employs the query network $g(.)$ and the key network $g'(.)$ to encode the query $q = g(Aug_q(x))$ and keys $k = g'(Aug_k(x))$. Similar to MoCo (He et al. 2020), the key network employs a momentum update using the query network. Additionally, we maintain a queue that stores the most recent key embeddings k and chronologically update the queue. This enables us to establish a contrastive embedding pool $A = B_q \cup B_k \cup queue$, where B_q and B_k represent vectorial embeddings corresponding to the query and key views, respectively. For each sample, the contrastive loss can be calculated by contrasting its query embedding with the remaining embeddings in pool A.

$$\begin{aligned} \mathcal{L}_{\text{cont}} (g; x, \tau, A) &= -\frac{1}{|P(x)|} \\ & \sum_{k_+ \in P(x)} \log \frac{\exp \left(q^\top k_+ / \tau \right)}{\sum_{k' \in A(x)} \exp \left(q^\top k' / \tau \right)}, \end{aligned} \tag{6}$$

where $A(x) = A \backslash \{q\}$ and $\tau \geq 0$ is the temperature parameter. Inspired by (Shen and Sanghavi 2019), DNNs first memorize the training data of easy-learning samples, then gradually adapt to noisy labels. We construct the positive set $P(x)$ with the predicted label from the Classifier (See Figure. 2). About the query view, we train the classifier f using cross-entropy loss,

$$\mathcal{L}_{\text{cls}} \left(f; x_i, c_{pse}^i \right) = \sum_{n=1}^{n_C} -s_n^i \log \left(f^n (x_i) \right), x_i \in X^T \tag{7}$$

where n_C indicates the number of categories in $\hat{C}_T$, n denotes the indices of labels, s_n^i denotes the n-th vector of c_{pse}^i, and f^n denotes the n-th output of f. Putting it all together, the overall loss function can be defined as $\mathcal{L} = \mathcal{L}_{\text{cls}} + \beta \mathcal{L}_{\text{cont}}$, where β is set as 0.01 to balance each loss component.

Methods	Office-Home					Office-31				VisDA+ImageCLEF-DA			
	Ar	Cl	Pr	Re	Avg.	A	D	W	Avg.	C	I	P	Avg.
DANN	68.97	53.37	79.70	82.09	71.03	83.43	81.36	84.22	83.00	76.25	64.75	59.75	66.92
RTN	68.72	59.97	77.04	86.00	72.93	86.70	88.64	83.22	86.19	81.50	67.75	62.75	70.67
OSBP	44.17	45.98	63.37	68.56	55.52	57.76	81.54	78.48	72.59	49.75	44.50	44.25	46.17
UAN	69.27	60.32	79.78	82.82	73.05	85.35	94.54	92.03	90.65	75.00	67.75	61.00	67.92
DCTN	64.77	42.09	65.25	70.11	60.56	88.84	89.18	83.73	87.25	66.25	55.75	50.50	57.50
MDAN	67.56	55.36	79.20	86.02	72.04	85.82	92.82	88.43	89.02	68.50	65.25	60.50	64.75
MDDA	44.66	34.54	54.93	53.24	46.84	84.91	89.16	89.60	87.89	60.25	44.50	36.50	47.08
UMAN	79.00	**64.68**	81.12	87.08	77.97	**90.22**	94.50	94.53	**93.08**	**88.00**	**83.25**	**70.50**	**80.58**
Data, Model Specifics: Available↑, Data, Model Specifics: Decentralized ↓													
HCLD^{2*}	**80.31**	62.27	**82.33**	**88.85**	78.42	80.00	96.55	**96.52**	91.02	77.25	67.54	60.05	68.28
HCLD^2	77.38	61.09	81.74	88.49	77.18	75.66	**97.22**	94.95	89.28	84.25	73.50	68.50	75.41

Table 1: Comparison with the State-Of-The-Art DA methods on three DA benchmarks (Backbone: Resnet-50) measured by Accuracy (%). The best numbers are highlighted in bold. The second numbers are highlighted with an underline. Different from HCLD^2, HCLD^{2*} implements HCLD^2 with the Pseudo-Soft-Label.
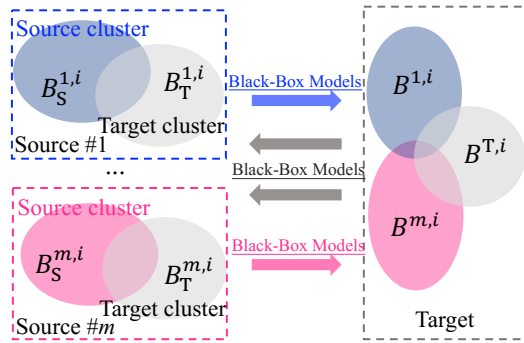

Figure 3: Overview of Mutual-Voting Decision (MVD).

MVD

Through the above HCLD strategy, we could optimize a model that is well-performing in shared classes and ambiguous in unknown classes. However, better adaptation performance depends on accurate inference for shared and unknown classes, which becomes challenging without multi-source data or parameters. Inspired by the consensus knowledge of shared classes among different domains, we consider utilizing cluster-level consensus from multi-source and target APIs to distinguish between shared and unknown classes. As the source and target APIs rarely misunderstand the non-existent category as the same shared class, we introduce an MVD strategy, which leverages the knowledge voting among the source and target views. Specifically, it calculates the voting scores in each class (the proportion of overlapping samples recognized as the same category in the dataset by all APIs, compared to the minimum number of all samples recognized as that category among these APIs) and calculates the mutual voting scores among the source and target views, which can be used to determine if it reaches a consensus. The overview of MVD is shown in Figure. 3. For the source view, given a pair of matched class clusters $B_S^{m,i}$ (obtain $B_S^{m,i}$ across the outputs of the m-th source model)

and $B_T^{m,i}$, we measure the cluster-level consensus via calculating the voting score $\{d_s^1, \ldots, d_s^{n_C}\}$,

$$d_s^{i,m} = \frac{(B_S^{m,i} \cap B_T^{m,i})}{\arg\min_{j \in \{1,\ldots,M\}}(B_S^{j,i}, B_T^{j,i})}, i \in \mathcal{C}_{S_m} \quad (8)$$

$$d_s^i = \arg \max_{j \in \{1,\ldots,M\}} d_s^{i,j} \quad (9)$$

Similarly, we calculate the voting score $\{d_t^1, \ldots, d_t^{n_c}\}$ in the target view. Then, the mutual-voting score of two views for each union source class can be calculated as:

$$\mathcal{S}_c = \frac{d_t^c + d_s^c}{2}, c \in \hat{\mathcal{C}}_T \quad (10)$$

For each $\mathcal{S}_c$, we can predict the class c with a validated threshold λ. This either assigns class c to one of the union source classes ($\mathcal{S}_c < \lambda$) or rejects it as an "unknown" class.

Experiments

Experimental Setup

Datasets. **Office-Home** (Venkateswara et al. 2017) is a DA benchmark that consists of four domains: Art (Ar), Clipart (Cl), Product (Pr), and Real World (Re). **Office-31** (Saenko et al. 2010) is another popular benchmark that consists of three domains: Amazon (A), Webcam (W), and Dslr (D). **VisDA2017+ImageCLEF-DA** is a combination of two datasets. VisDA2017 (Peng et al. 2018) is a DA dataset where the source domain contains simulated images (S) and the target domain contains real-world images (R). ImageCLEF-DA, on the other hand, is organized by selecting the common categories shared by three large-scale datasets: ImageCLEF (C), ImageNet (I), and Pascal VOC (P). Classes in the combined dataset are numbered as follows: Classes No. 1–7 represent the shared classes among the five datasets in alphabetical order. Classes No. 8–12 are the remaining classes from S and R domains. Classes No. 13–17 are the remaining classes from the C, I, and P domains. In UFDA, each domain contains two types of labels: shared and unknown. We use a matrix to describe the

specific UniMDA setting, called UMDA-Matrix (Yin et al. 2022), which is defined as $\begin{bmatrix} \frac{|\mathcal{C}_1|}{|\overline{\mathcal{C}}_{S_1}|} & ... & \frac{|\mathcal{C}_M|}{|\overline{\mathcal{C}}_{S_M}|} & \frac{|\mathcal{C}|}{|\overline{\mathcal{C}}_t|} \end{bmatrix}$. The first row is the size of the shared class of all the domains, and the second row denotes the unknown class. The first m columns are the label set of the multi-source domains, and the last one denotes the target domain. In this way, UniMDA settings can be determined by the division rule. To ensure a fair comparison with previous UniMDA works, we maintain the same UMDA-Matrix settings with UMAN.

Baseline Methods. The proposed HCLD^2 (HCLD & MVD) is compared with a range of State-Of-The-Art (SOTA) DA approaches. *i.e.*, including DANN (Ganin et al. 2016), RTN (Long et al. 2016), OSBP (Saito et al. 2018), MDAN (Zhao et al. 2018), MDDA (Zhao et al. 2020), UAN (You et al. 2019), DCTN (Xu et al. 2018), and UMAN (Yin et al. 2022). To ensure a fair comparison, we still the same evaluation metrics as those in the previous study (Yin et al. 2022), which represents the mean per-class accuracy over both the shared classes and the unknown class.

Since the UFDA setting is fairly new in this field, we also compare the other setting HCLD^{2*} based on the proposed HCLD^2. Different from HCLD^2, HCLD^{2*} implements HCLD^2 with the Pseudo-Soft-Label (PSL) which is generated by averaging the output of source models, we weigh each class by the number of source models containing this class.

Implementation Details. In UFDA, the architecture in each node can be either identical or radically different. However, to ensure a fair comparison with previous UniMDA works, we maintain a common model architecture. Specifically, we utilize ResNet-50 as the backbone for all tasks. The projection head of the contrastive network is a 2-layer MLPs that outputs 128-dimensional embeddings. For model optimization, we employ stochastic gradient descent (SGD) training with a momentum of 0.9. The learning rate is decayed using the cosine schedule, starting from a high value (*e.g.*, 0.005 for Office-31, Office-Home, and VisDA2017+ImageCLEF-DA) and decaying to zero. To follow the standard UniMDA training protocol, we use the same source and target samples, network architecture, learning rate, and batch size as in the UMAN (Yin et al. 2022). In decentralized training, the number of communication rounds r plays a crucial role. To ensure a fair comparison with traditional UniMDA works, we adopt $r = 1$ for all tasks. Furthermore, we implement all methods using PyTorch and conduct all experiments on an NVIDIA GeForce GTX 4*2080Ti, utilizing the default parameters for each method.

Experimental Results

Here we present the comparison between our method and the above baseline methods. Some results are directly chosen from (Feng et al. 2021). From the results in Table 1, despite the raw data and model specifics are not available, HCLD^2 still can perform comparably across almost all tasks compared with the traditional UniMDA setting. It also shows that, although HCLD^{2*}'s performance on

Methods				VisDA+ImageCLEF-DA			
PSL	PHL	GCLD	MVD	C	I	P	Avg.
✓				56.5	52.1	53.2	53.9
✓			✓	60.3	59.5	55.0	58.3
✓		✓		69.8	62.4	54.8	62.3
✓		✓	✓	77.3	67.5	60.1	68.3
	✓			44.3	43.0	37.6	41.6
	✓		✓	53.5	47.3	44.3	48.4
	✓	✓		74.5	70.0	55.2	66.6
	✓	✓	✓	84.3	73.5	68.5	75.4

Table 2: Ablation Study. PSL: Pseudo-Soft-Label. PHL: Pseudo-Hot-Label. GCLD: Gaussian Mixture Model-based Contrastive Label Disambiguation. MVD: Mutual-Voting Decision.

VisDA+ImageCLEF-DA is not ideal, it achieves SOTA results across several tasks on Office-Home and Office-31. The results highlight the efficacy of our proposed HCLD^2 again and demonstrate the instability of directly using the soft outputs for the pseudo-label generation.

Ablation Study

Overall Component Effectiveness. We study the effectiveness of three key components (PHL Generation, GCLD, and MVD) in HCLD^2, with results shown in Table 2. Results show that both GCLD and MVD significantly improved accuracy compared to the approach that removes MVD and GCLD only trains a classifier with the pseudo-labels (PHL or PSL). By combining these two components we can obtain the best performance. Suffer from the one-hot setting, the method exclusively trains a classifier employing the PHL, resulting in consistently lower accuracy compared to the PSL. However, intriguingly, the integration of GCLD yields a remarkable outcome where the PHL-based approach significantly outperforms the PSL-based approach by a substantial margin.

Effectiveness of the PHL Generation. To further analyze the impact of different pseudo-label generated methods, we report the performance of HCLD^{2*} and HCLD^2 with varying settings of category in Table 3. We can see that HCLD^{2*} works better than HCLD^2 when the intersection of the multi-source label sets is non-empty. However, when the intersection is empty, the performance of HCLD^{2*} will suddenly decline along with the accuracy of PSL. On the other hand, HCLD^2 performs well with all category settings and shows a more stable performance compared with HCLD^{2*}, which is sensitive to different category settings.

Effectiveness of GCLD. In Figure. 4a, we report the performance of PSL with and without GCLD, and PHL with GCLD. As illustrated, PSL with GCLD outperforms the approach without GCLD by a large margin. In the initial epochs, PHL with GCLD may suffer from the one-hot setting. As the number of training epochs increases, PHL with GCLD will surpass the performance of PSL with GCLD.

UMDA-Matrix	M	Ar	Cl	Pr	Re	Avg.
$\begin{bmatrix} 3 & 3 & 2 & 8 \\ 2 & 2 & 1 & 52 \end{bmatrix}$	P	36.4	47.3	55.7	66.8	51.5
	S	48.8	50.6	65.9	76.3	60.4
	H	68.9	51.9	72.6	80.5	68.5
$\begin{bmatrix} 4 & 3 & 3 & 10 \\ 2 & 2 & 2 & 49 \end{bmatrix}$	P	47.9	43.5	57.7	65.5	53.6
	S	59.9	51.0	68.2	76.2	63.9
	H	67.6	52.6	69.9	78.2	67.1
The intersection of MS: Empty ↑, Non-empty ↓						
$\begin{bmatrix} 4 & 4 & 4 & 10 \\ 2 & 2 & 2 & 50 \end{bmatrix}$	P	68.3	52.7	72.4	75.4	67.2
	S	80.3	62.3	82.3	88.9	78.4
	H	77.4	61.1	81.7	88.5	77.2
$\begin{bmatrix} 6 & 6 & 6 & 10 \\ 2 & 2 & 2 & 50 \end{bmatrix}$	P	75.0	55.9	73.9	82.8	71.9
	S	83.7	64.9	83.9	90.3	80.7
	H	81.4	62.1	80.9	89.2	78.4

Table 3: Comparison with different category settings on Office-Home measured by Accuracy (%). MS: Multi-Source label sets. M: Methods. P: Pseudo-Soft-Label. S: Implementation of HCLD^2 with P. H: Our proposed HCLD^2.

Methodologies	A	D	W	Avg.
w/o MVD	69.48	90.26	89.77	83.17
MVD (w/o target view)	71.53	94.52	92.43	86.16
MVD (w/o source view)	75.18	96.55	93.59	88.44
MVD	75.66	97.22	94.95	89.28

Table 4: Comparison with different shared-class decision strategy on Office-31 measured by Accuracy (%).

		FDA (r)				
	SFDA	0.2	0.5	1	5	10
A	69.59	72.76	75.09	75.66	77.39	77.72
D	86.92	92.3	95.04	97.22	97.27	97.11
W	90.59	92.49	94.34	94.95	95.54	96.93
Avg.	83.03	85.85	88.15	89.28	90.07	90.59

Table 5: Accuracies (%) on Office-31 for SFDA and FDA with various communication rounds r (per epoch).

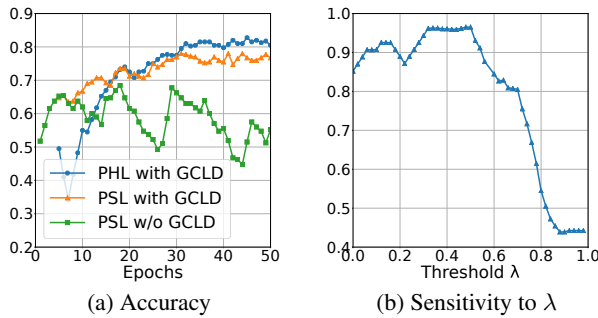(a) Accuracy (b) Sensitivity to λ

Figure 4: (a) indicates the Accuracy of Pseudo-labels on task C in VisDA+ImageCLEF-DA. (b) plots the sensitivity to parameter λ on task Dslr in Office-31. PSL: Pseudo-Soft-Label. PHL: Pseudo-Hot-Label. GCLD: Gaussian Mixture Model-based Contrastive Label Disambiguation. The vertical axes ((a) and (b)) all represent Accuracy (%).

Effectiveness of MVD. In Table 4, we show results for the single-view and mutual-voting decision strategies. As a baseline, we implement HCLD^2 without incorporating any shared-class decision strategy. We establish the shared classes through voting outcomes within the source or target node for the single view. As we can see, MVD yields the most favorable results compared to any other single-view approach, although every single strategy exhibits improved performance over the baseline. Moreover, we study the parameter λ on task Dslr. As shown in Figure. 4b, within a wide range of λ (0.3-0.5), the performance only varies to a small degree, showing that our method is robust to different choices of λ.

Why not Source-Free DA. Although FDA and SFDA are similar to some extent (e.g., only the pre-trained source model is accessible to the target domain), they are essentially different. FDA has an important assumption, i.e., the decentralized source clients keep communicating the updated source black-box models during the training process, whereas this does not hold in SFDA at all. Both our proposed scenario UFDA and our method HCLD^2 heavily rely on this assumption and aim to make the black-model communication in a more practical condition. Indeed, the difference between the SFDA and FDA settings of our method could be reflected in Table 5. As seen, without black-box model communication in SFDA, the performance of our model significantly drops. Moreover, the FDA performance increases with the communication round r.

Conclusion

This work investigated a more practical scenario, UFDA, where we relax the comprehensive assumptions such as configuration specifics nor the prior label set overlap across multi-source and target domains as in most FDA scenarios. We propose a new optimization methodology HCLD^2 to address UFDA and cluster-level strategy called MVD to distinguish shared and unknown classes during inference. Through extensive evaluations of three benchmark datasets, we demonstrate that HCLD^2 is capable of achieving comparable performance as conventional MDA baselines even with much less source knowledge. In the future, we may explore methods to further minimize additional assumptions (e.g., Source label sets) in our UDFA, aiming for a more relaxed FDA scenario.

Acknowledgments

References

Ahmed, S. M.; Raychaudhuri, D. S.; Paul, S.; Oymak, S.; and Roy-Chowdhury, A. K. 2021. Unsupervised multi-source domain adaptation without access to source data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10103–10112.

Ben-David, S.; Blitzer, J.; Crammer, K.; Kulesza, A.; Pereira, F.; and Vaughan, J. W. 2010. A theory of learning from different domains. *Machine learning*, 79: 151–175.

Blitzer, J.; Crammer, K.; Kulesza, A.; Pereira, F.; and Wortman, J. 2007. Learning bounds for domain adaptation. *Advances in neural information processing systems*, 20.

Chen, L.; Lou, Y.; He, J.; Bai, T.; and Deng, M. 2022a. Evidential neighborhood contrastive learning for universal domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 6258–6267.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, 1597–1607. PMLR.

Chen, Z.; Gu, S.; Lu, G.; and Xu, D. 2022b. Exploiting intra-slice and inter-slice redundancy for learning-based lossless volumetric image compression. *IEEE Transactions on Image Processing*, 31: 1697–1707.

Chen, Z.; Gu, S.; Zhu, F.; Xu, J.; and Zhao, R. 2021. Improving facial attribute recognition by group and graph learning. In *2021 IEEE International Conference on Multimedia and Expo (ICME)*, 1–6. IEEE.

Chen, Z.; Lu, G.; Hu, Z.; Liu, S.; Jiang, W.; and Xu, D. 2022c. LSVC: a learning-based stereo video compression framework. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6073–6082.

Chen, Z.; Zhou, J.; Wang, X.; Swanson, J.; Chen, F.; and Feng, D. 2017. Neural net-based and safety-oriented visual analytics for time-spatial data. In *2017 International Joint Conference on Neural Networks (IJCNN)*, 1133–1140. IEEE.

Csurka, G.; Hospedales, T. M.; Salzmann, M.; and Tommasi, T. 2022. Visual Domain Adaptation in the Deep Learning Era. *Synthesis Lectures on Computer Vision*, 11(1): 1–190.

Dong, J.; Fang, Z.; Liu, A.; Sun, G.; and Liu, T. 2021. Confident Anchor-Induced Multi-Source Free Domain Adaptation. *Advances in Neural Information Processing Systems*, 34.

Fantauzzo, L.; Fan, E.; Caldarola, D.; Tavera, A.; Cermelli, F.; Ciccone, M.; and Caputo, B. 2022. Feddrive: Generalizing federated learning to semantic segmentation in autonomous driving. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 11504–11511. IEEE.

Feng, H.; You, Z.; Chen, M.; Zhang, T.; Zhu, M.; Wu, F.; Wu, C.; and Chen, W. 2021. KD3A: Unsupervised Multi-Source Decentralized Domain Adaptation via Knowledge Distillation. In *ICML*, 3274–3283.

Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F.; Marchand, M.; and Lempitsky, V. 2016.

Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1): 2096–2030.

Gilad-Bachrach, R.; Dowlin, N.; Laine, K.; Lauter, K.; Naehrig, M.; and Wernsing, J. 2016. Cryptonets: Applying neural networks to encrypted data with high throughput and accuracy. In *International conference on machine learning*, 201–210. PMLR.

Guo, E.; Fu, H.; Zhou, L.; and Xu, D. 2023. Bridging Synthetic and Real Images: a Transferable and Multiple Consistency aided Fundus Image Enhancement Framework. *IEEE Transactions on Medical Imaging*.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9729–9738.

Hoffman, J.; Mohri, M.; and Zhang, N. 2018. Algorithms and theory for multiple-source adaptation. *Advances in Neural Information Processing Systems*, 31.

Khosla, P.; Teterwak, P.; Wang, C.; Sarna, A.; Tian, Y.; Isola, P.; Maschinot, A.; Liu, C.; and Krishnan, D. 2020. Supervised contrastive learning. *Advances in neural information processing systems*, 33: 18661–18673.

Kundu, J. N.; Venkat, N.; Babu, R. V.; et al. 2020. Universal source-free domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4544–4553.

Li, J.; Xiong, C.; and Hoi, S. C. 2021. Learning from noisy data with robust representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 9485–9494.

Li, S.; Xia, X.; Ge, S.; and Liu, T. 2022. Selective-supervised contrastive learning with noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 316–325.

Li, X.; Gu, Y.; Dvornek, N.; Staib, L. H.; Ventola, P.; and Duncan, J. S. 2020. Multi-site fMRI analysis using privacy-preserving federated learning and domain adaptation: ABIDE results. *Medical Image Analysis*, 65: 101765.

Liang, J.; Hu, D.; and Feng, J. 2020. Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In *International Conference on Machine Learning*, 6028–6039. PMLR.

Liang, J.; Hu, D.; Feng, J.; and He, R. 2022. Dine: Domain adaptation from single and multiple black-box predictors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8003–8013.

Liu, X.; Xi, W.; Bai, G.; Wang, Z.; Liu, Z.; and Zhao, J. 2022. M2N: Mutual constraint network for multi-level unsupervised domain adaptation. *Neurocomputing*, 487: 269–279.

Liu, X.; Xi, W.; Li, W.; Xu, D.; Bai, G.; and Zhao, J. 2023. Co-MDA: Federated Multi-source Domain Adaptation on Black-box Models. *IEEE Transactions on Circuits and Systems for Video Technology*, 1–1.

Long, M.; Zhu, H.; Wang, J.; and Jordan, M. I. 2016. Unsupervised domain adaptation with residual transfer networks. *Advances in neural information processing systems*, 29.

McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282. PMLR.

Mohassel, P.; and Rindal, P. 2018. ABY3: A mixed protocol framework for machine learning. In *Proceedings of the 2018 ACM SIGSAC conference on computer and communications security*, 35–52.

Mohassel, P.; and Zhang, Y. 2017. Secureml: A system for scalable privacy-preserving machine learning. In *2017 IEEE symposium on security and privacy (SP)*, 19–38. IEEE.

Peng, X.; Huang, Z.; Zhu, Y.; and Saenko, K. 2019. Federated Adversarial Domain Adaptation. In *International Conference on Learning Representations*.

Peng, X.; Usman, B.; Kaushik, N.; Wang, D.; Hoffman, J.; and Saenko, K. 2018. Visda: A synthetic-to-real benchmark for visual domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2021–2026.

Permuter, H.; Francos, J.; and Jermyn, I. 2006. A study of Gaussian mixture models of color and texture features for image classification and segmentation. *Pattern recognition*, 39(4): 695–706.

Qu, S.; Zou, T.; Röhrbein, F.; Lu, C.; Chen, G.; Tao, D.; and Jiang, C. 2023. Upcycling models under domain and category shift. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 20019–20028.

Saenko, K.; Kulis, B.; Fritz, M.; and Darrell, T. 2010. Adapting visual category models to new domains. In *European conference on computer vision*, 213–226. Springer.

Saito, K.; Kim, D.; Sclaroff, S.; and Saenko, K. 2020. Universal domain adaptation through self supervision. *Advances in neural information processing systems*, 33: 16282–16292.

Saito, K.; and Saenko, K. 2021. Ovanet: One-vs-all network for universal domain adaptation. In *Proceedings of the ieee/cvf international conference on computer vision*, 9000–9009.

Saito, K.; Yamamoto, S.; Ushiku, Y.; and Harada, T. 2018. Open set domain adaptation by backpropagation. In *Proceedings of the European conference on computer vision (ECCV)*, 153–168.

Shan, X.; Ma, T.; and Wen, Y. 2023. Prediction of common labels for universal domain adaptation. *Neural Networks*.

Shen, Y.; and Sanghavi, S. 2019. Learning with bad training data via iterative trimmed loss minimization. In *International Conference on Machine Learning*, 5739–5748. PMLR.

Shui, C.; Chen, Q.; Wen, J.; Zhou, F.; Gagné, C.; and Wang, B. 2022. A novel domain adaptation theory with Jensen–Shannon divergence. *Knowledge-Based Systems*, 257: 109808.

Shui, C.; Li, Z.; Li, J.; Gagné, C.; Ling, C. X.; and Wang, B. 2021. Aggregating from multiple target-shifted sources. In *International Conference on Machine Learning*, 9638–9648. PMLR.

Tian, J.; Zhang, J.; Li, W.; and Xu, D. 2022. VDM-DA: Virtual Domain Modeling for Source Data-Free Domain Adaptation. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(6): 3749–3760.

Venkateswara, H.; Eusebio, J.; Chakraborty, S.; and Panchanathan, S. 2017. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 5018–5027.

Voigt, P.; and Von dem Bussche, A. 2017. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, 10(3152676): 10–5555.

Wu, G.; and Gong, S. 2021. Collaborative optimization and aggregation for decentralized domain generalization and adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 6484–6493.

Wu, K.; Shi, Y.; Han, Y.; Shao, Y.; and Li, B. 2021. Blackbox Probe for Unsupervised Domain Adaptation without Model Transferring. *arXiv preprint arXiv:2107.10174*.

Wu, Z.; Xiong, Y.; Yu, S. X.; and Lin, D. 2018. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3733–3742.

Xu, R.; Chen, Z.; Zuo, W.; Yan, J.; and Lin, L. 2018. Deep cocktail network: Multi-source unsupervised domain adaptation with category shift. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3964–3973.

Yang, Q.; Liu, Y.; Chen, T.; and Tong, Y. 2019. Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10(2): 1–19.

Yin, Y.; Yang, Z.; Hu, H.; and Wu, X. 2022. Universal multi-Source domain adaptation for image classification. *Pattern Recognition*, 121: 108238.

You, K.; Long, M.; Cao, Z.; Wang, J.; and Jordan, M. I. 2019. Universal domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2720–2729.

Zhao, H.; Zhang, S.; Wu, G.; Moura, J. M.; Costeira, J. P.; and Gordon, G. J. 2018. Adversarial multiple source domain adaptation. *Advances in neural information processing systems*, 31.

Zhao, S.; Wang, G.; Zhang, S.; Gu, Y.; Li, Y.; Song, Z.; Xu, P.; Hu, R.; Chai, H.; and Keutzer, K. 2020. Multi-source distilling domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 12975–12983.

Zhao, Y.; Zhong, Z.; Luo, Z.; Lee, G. H.; and Sebe, N. 2022. Source-free open compound domain adaptation in semantic segmentation. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10): 7019–7032.

Zheltonozhskii, E.; Baskin, C.; Mendelson, A.; Bronstein, A. M.; and Litany, O. 2022. Contrast to divide: Self-supervised pre-training for learning with noisy labels. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1657–1667.