The following publication Y. Lu, H. Huang, W. K. Wong, X. Hu, Z. Lai and X. Li, "Adaptive Dispersal and Collaborative Clustering for Few-Shot Unsupervised Domain Adaptation" in IEEE Transactions on Image Processing, vol. 34, pp. 4273-4285, 2025 is available at https://doi.org/10.1109/TIP.2025.3581007.

Adaptive Dispersal and Collaborative Clustering for Few-shot Unsupervised Domain Adaptation

Yuwu Lu, *Senior Member, IEEE,* Haoyu Huang, *Member, IEEE,* Wai Keung Wong, Xue Hu, Zhihui Lai, *Member, IEEE,* and Xuelong Li, *Fellow, IEEE*

***Abstract*—Unsupervised domain adaptation is mainly focused on the tasks of transferring knowledge from a fully-labeled source domain to an unlabeled target domain. However, in some scenarios, the labeled data are expensive to collect, which cause an insufficient label issue in the source domain. To tackle this issue, some works have focused on few-shot unsupervised domain adaptation (FUDA), which transfers predictive models to an unlabeled target domain through a source domain that only contains a few labeled samples. Yet the relationship between labeled and unlabeled source domains are not well exploited in generating pseudo-labels. Additionally, the few-shot setting further prevents the transfer tasks as an excessive domain gap is introduced between the source and target domains. To address these issues, we newly proposed an adaptive dispersal and collaborative clustering (ADCC) method for FUDA. Specifically, for the shortage of the labeled source data, a collaborative clustering algorithm is constructed that expands the labeled source data to obtain more distribution information. Furthermore, to alleviate the negative impact of domain-irrelevant information, we construct an adaptive dispersal strategy that introduces an intermediate domain and pushes both the source and target domains to this intermediate domain. Extensive experiments on the Office31, Office-Home, miniDomainNet, and VisDA-2017 datasets showcase the superior performance of ADCC compared to the state-of-the-art FUDA methods.**

***Index Terms*—Unsupervised domain adaptation, Few-shot, Image classification, Collaborative clustering.**

I. Introduction

Deep neural networks (DNNs) have achieved significant success in various computer vision tasks, including image recognition [1], [2], semantic segmentation [3], [4], and object detection [5], [6]. However, the remarkable results of the DNNs-based models are largely dependent on the costly and time-consuming labeled data. When the distributions of the test data deviate from those of the training data, the DNNs-based models may suffer from performance degradation due to domain shift issue [7]–[9]. An appealing way to tackle this problem is unsupervised domain adaptation (UDA) [10]–[12], which transfers knowledge from a fully-labeled domain (i.e., source domain) to an unlabeled domain (i.e., target domain) and generalizes well in the target domain.

Existing studies on UDA are primarily based on the setting where the source domain contains fully labeled data, while the target domain contains unlabeled data. Additionally, the source and target domains are assumed to share the same label space. To ensure that the adaptive model can effectively learn knowledge from the source domain and transfer it to the target domain, researchers have been made great efforts, including learning transferable representation via diffusion model [13], [14], learning domain-invariant representations via dynamic network [10], [15], [16], and learning an adaptive classifier based on the distance metric minimization constrain [17], [18].

Although existing UDA works have significantly advance on the unlabeled target domain, the standard assumption of UDA is that the source domain contains sufficient labeled data to support model training. However, the scenes in a real-world scenario are complex and diverse, and the source domain may suffer from high annotation costs or limited labeled data. For example, the data annotation of medical images is expensive due to the high cost of inviting domain experts [19], [20]. Furthermore, the failure rate of rotating machines is very low, leading to rare labeled fault data in fault diagnosis [21], [22]. Hence, there is a strong motivation to achieve UDA with few labeled source data, commonly referred as few-shot unsupervised domain adaptation (FUDA) [23]–[26].

The challenges of FUDA mainly include three points: 1) only a few samples in the source domain are annotated, while the remaining source and target samples are unlabeled. 2) The effectiveness of domain adaptation can be compromised when transfer objects are insufficient, as this may lead to an inaccurate measurement of the domain gap, either overstating or understating the actual distribution shift between the source and target domains. 3) Previous UDA methods are not suitable for FUDA due to the insufficiency of the source knowledge. Recently, inspired by the self-supervised learning (SSL) strategy in cross-domain few-shot learning [27], [28], researchers have introduced SSL to explore the latent semantic correspondence between the source and target domains [23], [24]. For instance, CDS [23] considered the domain gap issue in SSL and proposed in-domain and across-domain

Manuscript received July 29, 2024; revised February 7, 2025 and June 1, 2025; accepted June 9, 2024. This work was supported in part by the National Natural Science Foundation of China under Grants 62176162 and 62076129, in part by the Guangdong Basic and Applied Basic Research Foundation under Grants 2023A1515012875 and 2022A1515140099. (*Corresponding author: Wai Keung Wong.*)

Yuwu Lu, Haoyu Huang, and Xue Hu are with the School of Artificial Intelligence, South China Normal University, Foshan 528225, China (e-mail: {luyuwu2008, hyhuang99, hx1430940232}@163.com).

Wai Keung Wong is with the Institute of Textiles and Clothing, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, and also with the Laboratory for Artificial Intelligence in Design, Hong Kong (e-mail: calvin.wong@polyu.edu.hk)

Zhihui Lai is with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518055, China (e-mail: lai_zhi_hui@163.com).

Xuelong Li is with the Institute of Artificial Intelligence (TeleAI), China Telecom Corporation Ltd, Beijing 100033, China. (e-mail: xuelong_li@ieee.org).

self-supervision to explore the discriminative features and domain-invariant representations. PCS [24] proposed a cross-domain prototypical learning strategy that enhances source representations and target task training.

Although previous methods made great progress in FUDA, most of the current FUDA methods do not fully consider the latent correspondence between labeled and unlabeled data of the source domain, which may introduce noise information during the adaptation process and aggravate the shortage of labeled samples. In particular, simply using limited labeled samples to guide pseudo-label assignment for unlabeled instances often overlooks the intrinsic semantic relations or distributional structures among source samples, leading to inaccurate or unstable predictions. Moreover, mislabeled source samples can propagate errors into subsequent feature alignment stages, further exacerbating negative transfer. This shortcoming becomes more severe when there is a large discrepancy between the labeled and unlabeled subsets, as the available labeled samples may not provide sufficient diversity to cover the global distribution of the source domain.

Besides, existing approaches generally align the source and target domains directly, but they overlook domain-irrelevant information in both domains. Such information—ranging from low-level background cues to spurious correlations embedded in certain categories—can distort the learned model representation when explicitly forced to match across domains. For instance, the labeled subset in the source domain might contain instances with background clutter or confounding factors unrelated to the class semantics, yet direct alignment treats these factors as equally important for cross-domain matching. Consequently, irrelevant features may be amplified and transferred to the target domain, compromising adaptability and leading to suboptimal performance. This can result in potential distortion in model training caused by irrelevant information in the source domain, which in turn may engender poor generalization to the target domain and ultimately cause negative transfer issues.

To solve the aforementioned challenges, in this paper, we propose a novel FUDA approach, adaptive dispersal and collaborative clustering (ADCC), for cross-domain image classification. Considering the latent correspondence between labeled data and unlabeled data in the source domain, we present an Intra-domain Collaborative Clustering (ICC) method that leverages the latent correspondence between labeled and unlabeled samples. By using labeled source data as factors of a dictionary, ICC jointly considers the feature similarities, distributional proximity, and semantic links between labeled and unlabeled instances. This collaborative process helps refine pseudo-labels for unlabeled source data by integrating multiple "clustering factors", effectively filtering out noise and improving the reliability of pseudo-label generation. Unlike conventional methods that either rely solely on cluster centers or treat all unlabeled data equally, our ICC strategy capitalizes on multiple labeled prototypes in each category to reduce the adverse impact of outliers and better capture subtle class boundaries.

Moreover, a cross-domain adaptive dispersal scheme is designed to alleviate the difficulty of domain alignment. Instead of forcing direct alignment between source and target distributions, which can amplify domain-irrelevant content, our method constructs an intermediate "mix-up" domain as a soft bridge between them. Specifically, in the early stage of training, we form a mix-up domain that is closer to the unlabeled target domain to encourage a more natural transition and reduce abrupt distribution shifts. Then, we adaptively adjust the mix-up ratio toward the source domain over time, while also drawing the target domain closer to this evolving intermediate distribution. This "curriculum-like" progression helps the model gradually adapt from target-like representations to source-like ones, thereby filtering out spurious or domain-irrelevant features. Effectively, label information from the source domain becomes more selectively transferred as the mix-up domain shifts, so that only key semantic patterns are consistently preserved, while extraneous or noisy attributes are de-emphasized.

Our main contributions can be summarized as follows:

- We propose a novel adaptive dispersal and collaborative clustering (ADCC) method for the challenging FUDA task, which further takes the relationship between labeled and unlabeled source data and domain-irrelevant information into consideration.
- We introduce Intra-domain Collaborative Clustering (ICC) to comprehensively explore the correspondence and distributional information among labeled and unlabeled source samples. By building a multi-factor dictionary, our model can generate more accurate pseudo-labels for unlabeled source samples, and thereby better combat the scarcity of labeled data.
- A novel cross-domain adaptive dispersal scheme is proposed to address the domain alignment problem in FUDA. By constructing an intermediate mix-up domain and gradually shifting focus from target-like to source-like distributions, our approach avoids the abrupt usage of domain-irrelevant features and achieves smoother and higher-fidelity transfer.
- Extensive experiments on four datasets were conducted to evaluate the performance of ADCC (i.e., Office31, Office-Home, miniDomainNet, and VisDA-2017) and it outperforms SOTA FUDA approaches with large margins.

The remainder of this article is structured as follows. Section II provides a brief review of related work in the field. The proposed ADCC method and its optimization algorithm are detailed in Section III. The experimental results and analysis for several FUDA tasks are presented in Section IV. Finally, we conclude this article in Section V.

II. Related Work

A. Unsupervised Domain Adaptation (UDA)

As a branch of transfer learning, UDA aims to transfer knowledge from a single source domain to a single target domain, under the strong assumption that the source domain has fully labeled data while the target domain lacks labels. Specifically, learning domain-invariant representations is the main objective of most UDA approaches. Based on the learning strategy for domain-invariant representations, current

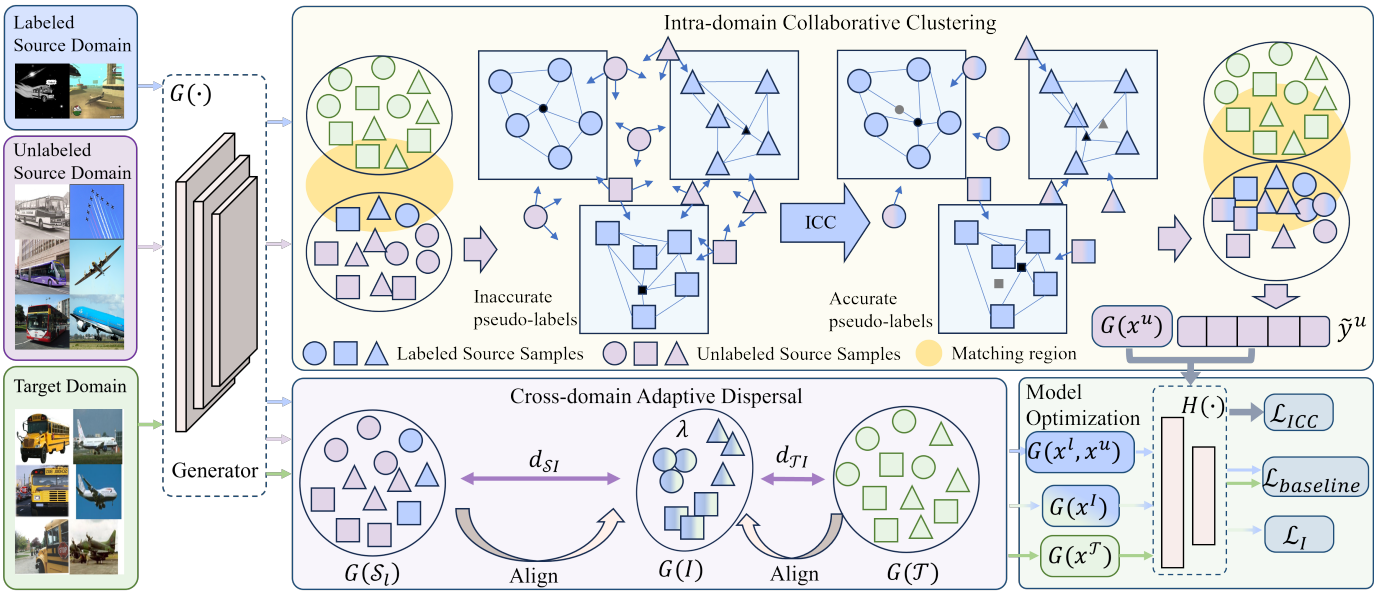


Fig. 1. The overview of the proposed ADCC method. In ICC, we first calculate the center for each cluster, then update the center by using multiple labeled source samples until the cluster center is not changed. Finally, the pseudo-labels generated by ICC are utilized to explore more discriminative information of the source domain. Specifically, the solid lines denote relationships between each sample and its corresponding cluster center (black and small shape). The arrows show how unlabeled source samples connect to their nearest cluster centers for pseudo-label generation, reflecting the collaborative representation process. In cross-domain adaptive dispersal, $d_{\mathcal{SI}}$ denotes the Wasserstein distance from the source domain to the intermediate domain and $d_{\mathcal{TI}}$ denotes the Wasserstein distance from the target domain to the intermediate domain, the adaptive mix-up ratio λ is updated by $d_{\mathcal{SI}}$ and $d_{\mathcal{TI}}$. By adaptively adjusting the mix-up ratio, the intermediate domain is gradually closer to the source domain, thus achieving smooth domain alignment. The matching region denotes the subset of source and target samples with overlapping feature distributions that facilitate effective domain alignment.

UDA approaches can be concretely divided into distance-based methods [17], [27]–[29] and adversarial learning-based methods [30]–[32]. The distance-based methods [17], [27]–[29] measure the domain discrepancy using a distance algorithm explicitly and minimize this discrepancy to learn domain-invariant representations. A typical work utilizing metric learning to achieve domain-invariant representations is JAN [29], which proposed the joint maximum mean discrepancy (JMMD) to simultaneously adjust the marginal and conditional distributions, creating robust feature representations for cross-domain classification. The adversarial learning-based methods [30]–[32] construct a min-max game between the feature generator and domain discriminator, confusing the discriminator to learn domain-invariant representations. DANN [33] proposed the gradient reversal layer (GRL) to perform the adversarial paradigm between the feature generator and domain discriminator.

Recently, researchers have focused on various factors in learning domain-invariant representations to further alleviate the domain shift issue. BuresNet [17] introduces the Conditional Kernel Bures (CKB) metric to measure and reduce the gap between conditional distributions, addressing common oversights in previous methods. The CKB metric, operating under the optimal transportation framework, can be integrated into deep networks as BuresNet, effectively extracting conditional invariant features for UDA. CAF [27] uses Wasserstein distance minimization for domain alignment and employs dual classifiers to enhance semantic consistency by aligning target features with source data class-wise. CDAN [30] proposed two conditioning strategies: multilinear conditioning to capture cross-covariance between features and classifier predictions, and entropy conditioning to control prediction uncertainty. DALN [31] introduces an adversarial learning network without a discriminator, repurposing the category classifier as the discriminator for explicit domain alignment and category distinction. Although previous methods have made significant progress in UDA tasks, most require sufficient labeled source data to support model training for image classification, which may not be suitable in the sparse source label setting of FUDA.

B. Few-shot Unsupervised Domain Adaptation (FUDA)

FUDA focuses on a more challenging setting where only a few (even one-shot) source domain samples are annotated, while the remaining source and target samples are unlabeled. Existing FUDA approaches [23]–[26], [34]–[36] mainly introduce self-supervised learning (SSL) to investigate a common embedding space across both the source and target domains. CDS [23] proposed in-domain SSL to learn discriminative features within a domain and across-domain SSL to better transfer knowledge in a few-shot source domain. PCS [24] utilized prototypes of the source and target domains to boost model performance and integrated all parts into an end-to-end framework. C-VisDiT [26] further considered the confidence of source samples and the transfer difficulty of the unlabeled target data in the FUDA problem. TSECS [34] learns high-level semantic features to improve image-to-class similarity measurement and employs a cross-domain self-training strategy and KL divergence minimization to better align source and target domains, enhancing classification performance. SAMix [35] uses a spectral sensitivity map and target-style image generation to enhance model generalization with limited unlabeled target domain samples. PAC [36] improves prototype

quality and enhances target domain generalization through new computation strategies for both in-domain and cross-domain contrastive learning.

Although significant progress has been made, current FUDA methods do not fully consider the correspondence between labeled and unlabeled samples in the source domain. Consequently, the model cannot obtain sufficient information to support knowledge transfer. Furthermore, the negative impact of domain-irrelevant information on domain alignment is not well mitigated in FUDA. Therefore, we propose the ADCC method to tackle these weaknesses.

C. Few-shot Learning (FSL)

The aim of FSL in image classification lies in building a trainable model capable of adapting to novel few-shot tasks. Current FSL methods can be roughly divided into transfer learning-based approaches [37], [38] and meta-learning-based methods [39]–[41]. In transfer learning-based methods, the model is first pre-trained on a large dataset containing base classes. Subsequently, the pre-trained model is fine-tuned for the few-shot tasks involving new categories. Ling et al. [37] proposed a pseudo-label acquisition scheme to alleviate the scarcity of labeled samples. In meta-learning-based methods, the model is trained to utilize only few-shot training data to perform well in new categories. Subramanyam et al. [39] constructed a contrastive distillation strategy to adjust the learned knowledge. Ji et al. [41] focused on catastrophic forgetting and overfitting problems in FSL, adopting the class-incremental learning technique to acquire knowledge in novel tasks. However, FSL cannot handle the unavailability of labels in the target domain. Therefore, FUDA facilitates the transfer of knowledge from the labeled source domain to the unlabeled target domain and addresses the issue of sparsely labeled samples in the source domain, making it more practical and challenging.

III. Methodology

To tackle the FUDA problem, we proposed a novel method named ADCC, which exploits the correspondence and distribution information from the few-shot source domain and constructs an intermediate mix-up domain to gradually align the source and target domains. For a global understanding, we provide the overview of ADCC in Figure 1.

A. Preliminary

In FUDA problem, we have two domains drawn from diverse distributions. The source domain is drawn from the source distribution $P_{\mathcal{S}}(x, y)$, but only a few data in the source domain are annotated, and the remaining source data are unlabeled. The labeled source data can be denoted as $\mathcal{S}_l = \{(x_i^l, y_i^l)\}_{i=1}^{n_{\mathcal{S}}^l}$ and the unlabeled source data can be denoted as $\mathcal{S}_u = \{(x_i^u)\}_{i=1}^{n_{\mathcal{S}}^u}$, $n_{\mathcal{S}}^l$ and $n_{\mathcal{S}}^u$ denote the number of labeled samples and unlabeled samples of source domain, respectively, and $n_{\mathcal{S}}^l \ll n_{\mathcal{S}}^u$. The unlabeled target domain $\mathcal{T} = \{(x_i^{\mathcal{T}})\}_{i=1}^{n_{\mathcal{T}}}$ contains $n_{\mathcal{T}}$ unlabeled samples drawn from the target distribution $P_{\mathcal{T}}(x, y)$. In this work, we focus on the image classification tasks, where both the source domain and target domain have the same label space $Y = \{1, 2, ..., K\}$ with K categories.

B. Intra-domain Collaborative Clustering

In FUDA, the main challenge of knowledge transfer is the shortage of labeled source data. Although there is a considerable amount of unlabeled source data, it often contains noise and domain-irrelevant information, which can degrade the learning of domain-invariant representations. In this context, domain-irrelevant features refer to information that do not contribute to effective domain alignment and may introduce noise or ambiguity, hindering adaptation. For example, on the Office-Home dataset, current SOTA approaches PCS [24] and C-VisDiT [26] only achieve 45.5% and 56.9% on the unlabeled source domain, respectively, demonstrating that simply relying on limited labeled data and naive handling of unlabeled data cannot yield sufficient discriminative information. Consequently, negative transfer becomes inevitable since domain-irrelevant parts are carried over to the target domain.

Although prototypical contrastive learning (e.g., PCS [24]) captures semantic structures based on distances between samples and their cluster centers, it overlooks the relationships between labeled and unlabeled source samples. In fact, both labeled $\mathcal{S}_l$ and unlabeled $\mathcal{S}_u$ source data share a common distribution $P_{\mathcal{S}}(x, y)$, indicating that they may contain highly similar semantic structures. Therefore, relying solely on cluster centers risks missing crucial label–unlabeled relations and can leave noise unfiltered.

To address these issues, we propose intra-domain collaborative clustering (ICC). By collaboratively representing unlabeled samples in terms of both labeled data and cluster structures, ICC makes more effective use of the labeled–unlabeled relationships to improve pseudo-label quality. This process not only filters out noisy and domain-irrelevant information in the unlabeled source domain but also furnishes richer supervision for learning. Consequently, ICC provides a stronger, domain-relevant signal for transfer and helps alleviate negative transfer across the source and target domains.

Specifically, we set multiple factors for each cluster based on the embedded representation of labeled source data and utilize collaboration representation-based distance measurement to calculate the pseudo-labels of $\mathcal{S}_u$. For the setting of FUDA, there are M-shot labeled source samples in every category, and the clusters are expressed as $C = \{C_k\}_{k=1}^{K}$. Thus, the clustering center of k-th cluster C_k is initialized by the mean of the labeled source samples as:

$$c_k = \frac{1}{M} \sum_{j=1}^{M} G(x_{j,k}^l), \tag{1}$$

where c_k denotes the clustering center of k-th cluster C_k, $G(\cdot)$ denotes the feature generator that generates the embedded representation of the samples, $x_{j,k}^l$ denotes the j-th labeled sample that belongs to the k-th category/cluster. Hence, for the k-th cluster, the factors can be defined as $F_k = [G(x_{1,k}^l), \ldots, G(x_{M,k}^l), c_k]$. Considering the relation between labeled source data and unlabeled source data, we introduce collaborative representation [42] to utilize multiple factors for representing the unlabeled data while measuring the distance between the labeled data and unlabeled data. For the

Algorithm 1 Intra-domain Collaborative Clustering

Input: All labeled source domain $\mathcal{S}_l = \{(x_i^l, y_i^l)\}_{i=1}^{n_S^l}$; a mini-batch of unlabeled source domain $\mathcal{S}_u = \{(x_j^u)\}_{j=1}^{B}$; batch-size B.
Output: Pseudo-labels P of samples in unlabeled source domain $P = \{p_j, \tilde{y}_j\}_{j=1}^{B}$.

```
Calculate the initialize cluster center by Eq. (1).
while true do
    for i in 1:I do
        Estimate the distance d_ik by Eq. (2).
        Locate the cluster label arg min_{1≤k≤K} d_ik.
        Update each of the cluster and their relevant center by Eq. (5).
    end for
    if The centers are not further change then
        Compute the soft pseudo-labels and hard pseudo-labels by Eq. (6).
        break
    end if
end while
```

i-th unlabeled source sample, the distance to the k-th category can be calculated as:

$$d_{i,k} = \|G(x_i^u - F_k\rho_{i,k})\|_2^2, \tag{2}$$

where $\rho_{i,k} \in \mathbb{R}^{M+1}$ denotes the sub-coefficient vector linked to the k-th cluster and $\rho_i = [\rho_{i,1}, \rho_{i,2}, ..., \rho_{i,K}]^T$. Following [42], $\rho_{i,k}$ can be analytically derived as:

$$\rho_{i,k} = (F_k^T F_k + \lambda \boldsymbol{I})^{-1} F_k^T \cdot G(x_i^u), \tag{3}$$

where $\boldsymbol{I}$ is a unit diagonal matrix. $\lambda = 0.01$ is the regularization parameter. Thus, the cluster result of the unlabeled sample x_i^u is $r_i = \text{argmin}_{1\le k\le K} d_{i,k}$. Then, each cluster and their centers can be updated as follows:

$$\hat{C}_{r_i} = C_{r_i} \cup G(x_i^u), \tag{4}$$

$$\hat{c}_j = \frac{1}{|\hat{C}_j|} \sum_{G(x_n)\in\hat{C}_j} G(x_n) \quad \forall j = 1, 2, ..., K, \tag{5}$$

where $\hat{c}_j$ is the updated center of the j-th cluster, which is used to update the factors in the next iteration.

When the cluster centers do not change after several iterations, we obtain the final cluster $C = \{C_1, C_2, ..., C_K\}$. The soft pseudo-label $p_i = [p_{i,1}, p_{i,2}, ..., p_{i,K}]$ of unlabeled source sample x_i^u can be obtained by:

$$p_{i,j} = \frac{e^{-\log(d_{i,j}/\tau)}}{\sum_{k=1}^{K} e^{-\log(d_{i,k}/\tau)}}, \tag{6}$$

where τ is a temperature parameter and $p_{i,j}$ denotes the predict probability that sample x_i^u belongs to the j-th category. Then, the hard label is $\tilde{y}_i = \text{argmax}_j p_{i,j}$. The concrete steps of our collaborative clustering are shown in Algorithm 1.

By considering the feature distribution and semantic structure of labeled samples, the clusters in our method are more accurate and robust. Thus, we utilize the hard pseudo-labels of $\mathcal{S}_u$ to learn more discriminative information from unlabeled source domain. Specifically, to mitigate the impact of the high uncertainty samples, we measure the entropy of the predicted probability distribution of a sample and exclude the high uncertainty samples in discriminative information learning. We utilize the cross-entropy loss to learn the discriminative information from high-confidence unlabeled source data:

$$\mathcal{L}_{ICC} = \sum_{\mathcal{H}(p_i)<\gamma} \mathcal{L}_{CE}(H(G(x_i^u)), \tilde{y}_i), \tag{7}$$

where $\mathcal{H}(\cdot)$ denotes the prediction entropy and $H(\cdot)$ denotes the classifier. $\gamma = 0.6$ is the entropy threshold parameter used to filter out samples containing a high rate of domain-irrelevant information. p_i is the soft pseudo-label corresponding to the unlabeled source sample x_i^u.

C. Cross-domain Adaptive Dispersal

Domain alignment in FUDA is inherently challenging due to the scarcity of labeled source data and the noise in unlabeled source samples. While previous methods [23], [24] utilized self-supervised prototypical learning to mine domain-invariant and transferable features, they often rely on explicit alignment between source and target domains. However, this rigid alignment can unintentionally introduce domain-irrelevant information, especially when dealing with unlabeled or noisy data, thereby impairing performance on the target task. Thus, we propose a cross-domain adaptive feature dispersal strategy. Unlike static alignment schemes, our strategy dynamically adjusts mix-up ratios between domains in a curricula-like manner. By initially aligning the mixed domain closer to the target domain and progressively leveraging the strong supervision from the source domain, the model can gradually bridge the domain gap while suppressing irrelevant features. This dynamic process ensures that only the most transferable and domain-relevant information is emphasized throughout training, leading to more robust and effective adaptation in the FUDA setting.

1) Intermediate Mix-up Domain Constructing: Based on the insight that the knowledge transfer is from the source domain to the target domain, we leverage the labeled source samples that are closer to the target samples, to construct the intermediate mix-up domain. The Euclidean distance is employed to estimate the similarity between the source and target samples. For the target-source sample pair $(x_i^{\mathcal{T}}, x_j^l)$ in each iteration, the similarity score can be defined as:

$$s_{ij} = \|G(x_i^{\mathcal{T}}) - G(x_j^l)\|_2. \tag{8}$$

Then, the index $\tilde{j}_i$ of the nearest labeled source sample for the target sample $x_i^{\mathcal{T}}$ can be obtained by the similarity score:

$$\tilde{j}_i = \arg\min_j(s_{ij}). \tag{9}$$

Therefore, the reformulated labeled source data $(\tilde{x}_i^l, \tilde{y}_i^l)$, obtained using the index $\tilde{j}_i$, can be expressed as:

$$(\tilde{x}_i^l, \tilde{y}_i^l) = (x_{\tilde{j}_i}^l, y_{\tilde{j}_i}^l). \tag{10}$$

As the mix-up technique has made great progress in self-supervised learning, we try to apply mix-up technique to our

FUDA transfer task. The motivation for utilizing the mix-up technique in FUDA is to represent the unlabeled target data with the nearest labeled source data, which matches the basic concept of adaptation. Therefore, we construct an intermediate mix-up domain $\mathcal{I} = \{(x_i^{\mathcal{I}}, y_i^{\mathcal{I}})\}_{i=1}^{n_{\mathcal{T}}}$ by the unlabeled target data and their relevant nearest labeled source data as:

$$x_i^{\mathcal{I}} = \lambda \tilde{x}_i^l + (1-\lambda) x_i^{\mathcal{T}}, \tag{11}$$

$$y_i^{\mathcal{I}} = \lambda \tilde{y}_i^l + (1-\lambda) y_i^{\mathcal{T}}, \tag{12}$$

where λ is the adaptive mix-up ratio. During training, the mix-up ratio can be adjusted to control the intermediate mix-up domain closer to the source domain or the target domain.

2) Adaptive Mix-up Ratio: The mix-up ratio in previous works was generally selected by the fixed or randomized sampling strategy, such as beta distribution and Dirichlet distribution [26], [43]–[45]. However, in FUDA setting, domain alignment becomes challenging due to the scarcity of transferable objectives in the few-shot source domain. Therefore, the key insight lies in the sampling strategy of the mix-up ratio. In addition, Mai et. al. [46] also claimed that if the mix-up ratio was sampled from some distribution or a fixed value, the feature alignment would still be difficult or lead to a suboptimal solution.

To remedy such an issue, we propose an adaptive mix-up ratio strategy to enhance feature alignment in FUDA. Specifically, we utilize the Wasserstein distance $d_w(\cdot,\cdot)$ to measure the distribution distance between the embedded representations of labeled source data $\mathcal{S}_l$, unlabeled target data $\mathcal{T}$, and intermediate mix-up domain data $\mathcal{I}$. The distance from $\mathcal{S}_l$ to $\mathcal{I}$ can be defined as $d_{\mathcal{SI}} = d_w(G(\mathcal{S}_l), G(\mathcal{I}))$ and the distance from $\mathcal{T}$ to $\mathcal{I}$ can be defined as $d_{\mathcal{TI}} = d_w(G(\mathcal{T}), G(\mathcal{I}))$. Compared to the unlabeled target domain, the labeled source domain contains more supervised information that is utilized to align features. Thus, we set the intermediate mix-up domain closer to the unlabeled target domain by a smaller initialize mix-up ratio in the early stage of model training. As the training goes on, the intermediate mix-up domain gradually gets closer to the source domain by adaptively increasing the mix-up ratio. Guided by the intermediate mix-up domain, the target is also pushed towards the source domain since the distance $d_{\mathcal{TI}}$ is already small. Hence, by adjusting the mix-up ratio adaptively in a curriculum learning way [47], the domain alignment in FUDA goes more smoothly, and the model is more robust in domain-irrelevant information.

Specifically, we first calculate a weighted factor v to estimate the degree of closeness from the labeled source domain to the intermediate mix-up domain:

$$v = \exp\left(-\frac{d_w(G(\mathcal{S}_l), G(\mathcal{I}))}{(d_w(G(\mathcal{S}_l), G(\mathcal{I})) + d_w(G(\mathcal{T}), G(\mathcal{I})))T}\right), \tag{13}$$

where T denotes the temperature factor and set to 0.05. During the adaptation process, v is small in the early stages since the domain $\mathcal{I}$ is initially set closer to the target domain. Then, we adjust the mix-up ratio λ by a linearly incremental component to gradually guide the intermediate mix-up domain closer to the source domain. The adaptive adjustment strategy for mix-up ratio λ can be defined as:

$$\lambda_e = \frac{e(1-v)}{E} + v\lambda_{e-1}, \tag{14}$$

where e is the current epoch number and E is the total number of the training epochs. To make sure that the stability of the mix-up ratio in training, we introduce a random perturbation on λ_e as:

$$\tilde{\lambda}_e = \mathrm{R}(U(\lambda_e - \sigma, \lambda_e + \sigma), \min = 0.0, \max = 1.0), \tag{15}$$

where σ is the random perturbation parameter and U is the relevant uniform distribution. The range of the mix-up ratio is restricted to [0.0,1.0] by $\mathrm{R}(\cdot,\cdot,\cdot)$.

3) Cross-domain Alignment.: Based on the adaptive mix-up strategy, alignment between the source and target domains can occur in a smoother way instead of directly aligning two distributions. Hence, we opt to directly utilize the cross-entropy loss for optimizing the intermediate mix-up domain:

$$\mathcal{L}_{\mathcal{I}} = \sum_{i=1}^{n_{\mathcal{T}}} \mathcal{L}_{CE}(H(G(x_i^{\mathcal{I}})), y_i^{\mathcal{I}}). \tag{16}$$

Algorithm 2 ADCC method for FUDA

Input: All labeled source domain data $\mathcal{S}_l = \{(x_i^l, y_i^l)\}_{i=1}^{n_{\mathcal{S}}^l}$; unlabeled source domain data $\mathcal{S}_u = \{(x_j^u)\}_{j=1}^{B}$; unlabeled target domain data; hyperparameters λ_{ICC} and $\lambda_{\mathcal{I}}$; batch-size B; maximum iteration number I; initial mix-up ratio λ.

Output: Predicted target domain labels $\{(\hat{y}_i^{\mathcal{T}})\}_{i=1}^{n_{\mathcal{T}}}$.

for i in 1:I **do**
- Randomly sample a mini-batch data from each of the unlabeled source and unlabeled target domains.
- Construct the intermediate domain $\mathcal{I}$ by the mix-up ratio λ and Eqs. (11) and (12).
- Align source and target domains to the intermediate domain through $\mathcal{L}_{\mathcal{I}}$.
- Calculate the pseudo-labels of unlabeled source domain data by Algorithm 1.
- Calculate the intra-domain collaborative clustering loss through $\mathcal{L}_{ICC}$.
- Calculate the baseline model loss function through $\mathcal{L}_{baseline}$.
- Calculate the final loss function by $\mathcal{L}_{ADCC}$.
- Update the mix-up ratio λ through Eqs. (13), (14), and (15).

end for

D. Baseline Model

In our work, we introduce a general FUDA model as the baseline model of our method, which is also constructed by a shared feature generator $G(\cdot)$ and a category classifier $H(\cdot)$. For labeled data in the source domain $\mathcal{S}_l$, we utilize cross-entropy loss to optimize the feature generator and category classifier for classification:

$$\mathcal{L}_{cls} = \sum_{i=1}^{n_{\mathcal{S}}^l} \mathcal{L}_{CE}(H(G(x_i^l)), y_i^l). \tag{17}$$

(a) Office31 (b) Office-Home (c) miniDomainNet (d) VisDA-2017

Fig. 2. Some image examples from four domain adaptation datasets.

For the unlabeled target data, following [32], the mutual information maximization loss is employed to improve the quality of the target pseudo-labels. The expression of the mutual information maximization loss is:

$$\mathcal{L}_{MI} = -\sum_{k=1}^{K} \hat{p}^{(k)} \log \hat{p}^{(k)} + \frac{1}{n_{\mathcal{T}}} \sum_{i=1}^{n_{\mathcal{T}}} \left\langle p_i^{\mathcal{T}}, \log p_i^{\mathcal{T}} \right\rangle, \quad (18)$$

where $p_i^{\mathcal{T}}$ is the prediction of sample $x_i^{\mathcal{T}}$, which generated by the classifier with a softmax layer. $\hat{p}^{(k)}$ is the k-th element of the mean of all samples' prediction and $\hat{p} = \frac{1}{n_{\mathcal{T}}} \sum_{i=1}^{n_{\mathcal{T}}} p_i^{\mathcal{T}}$. $\langle \cdot, \cdot \rangle$ denotes the inner product operation.

To facilitate cross-domain feature alignment, we additionally incorporate the prototypical self-supervised learning loss [24] to bolster the model's performance. Due to the page limit, more details of $\mathcal{L}_{SSL}$ can be referred to [24].

Hence, the comprehensive loss function of the baseline model can be formulated as:

$$\mathcal{L}_{baseline} = \mathcal{L}_{cls} + \lambda_{MI}\mathcal{L}_{MI} + \lambda_{SSL}\mathcal{L}_{SSL}, \quad (19)$$

where λ_{MI} and λ_{SSL} are hyper-parameters.

E. Learning for FUDA

Our ADCC method mainly constructs intra-domain collaborative clustering and cross-domain adaptive mix-up strategies for the FUDA problem. Finally, when integrated with the baseline model, the overall loss function with hyperparameters λ_{ICC} and $\lambda_{\mathcal{I}}$ of ADCC is expressed as:

$$\mathcal{L}_{ADCC} = \mathcal{L}_{baseline} + \lambda_{ICC}\mathcal{L}_{ICC} + \lambda_{\mathcal{I}}\mathcal{L}_{\mathcal{I}}. \quad (20)$$

The concrete steps of ADCC are shown in Algorithm 2.

IV. Experiments

A. Experimental Setup

1) Datasets: We evaluated and compared the SOTA approaches with our method on four popular datasets (i.e., **Office31, Office-Home, miniDomainNet,** and **VisDA-2017**). Sample images from these four datasets are shown in Figure 2 and the statics of all the datasets are shown in Table I.

The **Office31** [50] dataset is a popular domain adaptation benchmark that comprises 4,110 images spanning 31 categories across 3 domains: A (Amazon, the image are sampled from the online merchants), D (Dslr, the images are sampled from the high-resolution digital SLR camera), and W (Webcam, the images are sampled from the low-resolution web camera).

TABLE I
STATICS OF THE DATASETS USED TO EVALUATE THE PROPOSED METHOD

Dataset	Sub-domain	Abbr.	#Source sample	#Class
Office31	Amazon Dslr Webcam	A D W	1shot/3shots	31
Office-Home	Art Clipart Product Real world	Ar Cl Pr Rw	3%/6% source samples	65
miniDomainNet	Clipart Painting Real Sketch	Cli Pnt Rel Skt	1shot/3shots	126
VisDA-2017	Synthetic Real	S R	0.1%/1% source samples	12

The **Office-Home** [51] dataset is a more challenging domain adaptation benchmark, which consists of 15,500 images, covering 65 common categories distributed across 4 domains: Ar (Art, includes a more abstract and varied visual style images), Cl (Clipart, includes simplistic, cartoon-like images that pose a unique challenge due to their lack of texture and realistic detail), Pr (Product, comprises images of various consumer goods, typically presented in a clean and isolated manner), and Rw (Real-world, contains images captured from everyday environments, providing a realistic and diverse set of visual representations).

The **miniDomainNet** [52] dataset is a subset of the DomainNet dataset [53], featuring 126 common categories across 4 domains: Cli (Clipart, consists of simple, cartoon-like images with minimalistic design and lack of texture), Pnt (Painting, includes artistic representations that vary in style, color, and texture), Rel (Real-world, encompasses photographs of real objects and scenes), and Skt (Sketch, comprises hand-drawn images, which are typically black-and-white line drawings that lack the detailed texture and color information).

The **VisDA-2017** dataset [54] is a large-scale dataset comprising a source training set with 152,409 synthetic images and a target set with 55,400 real images. Both the source and

TABLE II
CLASSIFICATION ACCURACY (%) ON THE OFFICE31 DATASET FOR FUDA USING THE RESNET-50 BACKBONE. THE BEST RESULTS ARE BOLDED.

Method	Office31: Acc. (%) on the target domain (1-shot/3-shots)						Avg.
	A→D	A→W	D→A	D→W	W→A	W→D	
Source Only	33.6/51.2	21.5/44.8	43.5/57.4	60.8/82.9	41.5/52.3	63.3/83.2	44.0/62.0
MME [48] (ICCV'19)	21.5/51.0	12.2/54.6	23.1/60.2	60.9/89.7	14.0/52.3	62.4/91.4	32.3/66.5
CDAN [30] (CVPR'18)	11.2/43.7	6.2/50.1	9.1/65.1	54.8/91.6	10.4/57.0	41.6/89.8	22.2/66.2
MDDIA [49] (ICML'20)	45.0/62.9	54.5/65.4	55.6/67.9	84.4/93.3	53.4/70.3	79.5/93.2	62.1/75.5
DALN [31] (CVPR'22)	55.2/68.6	59.1/70.8	60.3/73.9	89.2/94.0	58.5/72.8	85.7/94.5	68.0/79.1
DGWA [10] (TMM'24)	58.7/69.9	62.3/71.7	64.1/75.6	89.9/94.1	61.3/74.2	89.4/95.2	70.9/80.1
CDS [23] (arXiv'20)	52.6/65.1	55.2/68.8	65.7/71.2	76.6/88.1	59.7/71.0	73.3/87.3	63.9/75.3
PCS [24] (CVPR'21)	60.2/78.2	69.8/82.9	76.1/76.4	90.6/94.1	71.2/76.3	91.8/96.0	76.6/84.0
C-VisDiT [26] (ICCV'23)	74.1/82.7	72.3/86.0	75.7/76.5	93.2/95.0	76.4/76.9	94.2/97.0	81.0/85.7
ADCC	**76.7/86.1**	**74.8/89.7**	**78.1/79.3**	**94.7/96.3**	**79.8/81.0**	**95.9/98.3**	**83.0/89.0**

TABLE III
CLASSIFICATION ACCURACY (%) ON THE OFFICE-HOME DATASET FOR FUDA USING THE RESNET-50 BACKBONE. THE BEST RESULTS ARE BOLDED.

Method	Office-Home: Acc. (%) on the target domain												Avg.
	Ar→Cl	Ar→Pr	Ar→Rw	Cl→Ar	Cl→Pr	Cl→Rw	Pr→Ar	Pr→Cl	Pr→Rw	Rw→Ar	Rw→Cl	Rw→Pr	
3% labeled source sample													
Source Only	24.8	39.1	43.4	22.7	31.5	30.2	29.7	28.2	40.9	35.8	22.6	44.6	32.8
MME [48] (ICCV'19)	4.5	15.4	25.0	28.7	34.1	37.0	25.6	25.4	44.9	39.3	29.0	52.0	30.1
CDAN [30] (CVPR'18)	5.0	8.4	11.8	20.6	26.1	27.5	26.6	27.0	40.3	38.7	25.5	44.9	25.2
MDDIA [49] (ICML'20)	21.7	37.3	42.8	29.4	43.9	44.2	37.7	29.5	51.0	47.1	29.2	56.4	39.1
DALN [31] (CVPR'22)	26.8	45.5	52.2	36.8	46.1	50.7	49.9	47.5	63.3	60.0	35.2	67.5	48.5
DGWA [10] (TMM'24)	28.2	46.9	54.3	44.6	47.8	51.2	54.2	41.5	65.2	63.9	40.7	69.1	50.6
CDS [23] (arXiv'20)	43.8	55.5	60.2	51.5	56.4	59.6	51.3	46.4	64.5	62.2	52.4	70.2	56.2
PCS [24] (CVPR'21)	42.1	61.5	63.9	52.3	61.5	61.4	58.0	47.6	73.9	66.0	52.5	75.6	59.7
C-VisDiT [26] (ICCV'23)	44.1	66.8	67.0	54.9	66.4	66.8	60.5	47.9	75.7	67.2	51.6	78.8	62.3
ADCC	**45.3**	**70.2**	**68.2**	**56.3**	**68.7**	**72.2**	**61.7**	**48.1**	**76.3**	**68.6**	**53.9**	**80.1**	**64.1**
6% labeled source sample													
Source Only	28.7	44.2	47.9	31.4	42.5	39.4	37.2	33.3	59.7	47.8	34.5	61.4	42.3
MME [48] (ICCV'19)	27.6	43.2	49.5	41.1	46.6	49.5	43.7	30.5	61.3	54.9	37.3	66.8	46.0
CDAN [30] (CVPR'18)	26.2	33.7	44.5	34.8	42.9	44.7	42.9	36.0	59.3	54.9	40.1	63.6	43.6
MDDIA [49] (ICML'20)	25.1	44.5	51.9	35.6	45.7	50.3	48.3	37.1	64.5	58.2	36.9	68.4	47.2
DALN [31] (CVPR'22)	30.3	48.7	55.6	40.3	49.3	53.8	53.7	41.3	67.6	61.3	39.6	70.1	51.0
DGWA [10] (TMM'24)	30.7	50.1	57.5	47.2	51.3	55.5	56.7	44.5	68.7	65.7	43.6	71.3	53.6
CDS [23] (arXiv'20)	45.4	60.4	65.5	54.9	59.2	63.8	55.4	49.0	71.6	66.6	54.1	75.4	60.1
PCS [24] (CVPR'21)	46.1	65.7	69.2	57.1	64.7	66.2	61.4	47.9	75.2	67.0	53.9	76.6	62.6
C-VisDiT [26] (ICCV'23)	46.5	69.8	72.5	60.2	71.2	71.2	64.1	49.0	78.5	69.1	52.8	80.1	65.4
ADCC	**47.3**	**72.1**	**73.2**	**64.8**	**73.7**	**71.7**	**65.8**	**50.3**	**78.9**	**70.0**	**54.2**	**80.9**	**67.0**

target sets share 12 common categories.

2) Protocols: Due to the challenges inherent in the FUDA setting, the standard protocols used in the above four datasets are not suitable for comparison. In line with prior research [23], [24], we conducted experiments utilizing 1-shot and 3-shot source labels per class on the Office31 and miniDomainNet datasets. For the Office-Home dataset, we randomly sampled 3% and 6% labeled source samples for every single class following [23]. For VisDA-2017, we show results in settings with 0.1% and 1% labeled source domain as suggested in [26]. Such sampling was conducted for five times and the average result is reported.

3) Implementation Details: We implemented and evaluated our method on the PyTorch platform [55]. The version of PyTorch is 1.8.0. All the experiments were run on a Nvidia GeForce RTX-4090 GPU, and the CUDA version is 12.0. For the miniDomainNet and VisDA-2017 datasets, we employ ResNet-101 [56] as the backbone network. For the Office31 and Office-Home datasets, we adopt ResNet-50 as the backbone network. In both ResNet-50 and ResNet-101, the last fully connected (FC) layer is substituted with a 512-dimensional linear layer. We utilize stochastic gradient descent (SGD) as the optimizer, with a momentum parameter of 0.9 and a weight decay of 1e-3. Throughout training, the learning rate was initially set to 1e-3 and adjusted using the LambdaLR strategy [55].

4) Competitors: We selected several state-of-the-art (SOTA) approaches to evaluate the advancements of the proposed ADCC method. These include FUDA SOTA approaches such as CDS [23], PCS [24], and C-VisDiT [26]. Additionally, to highlight the challenges of the FUDA scenario, we selected several standard domain adaptation approaches and reproduced them under the FUDA setting. These include MME [48], CDAN [30], DALN [31], DGWA

TABLE IV
CLASSIFICATION ACCURACY (%) ON THE MINIDOMAINNET AND VISDA-2017 DATASET FOR FUDA USING THE RESNET-101 BACKBONE. THE BEST RESULTS ARE BOLDED.

Method	miniDomainNet: Acc. (%) on the target domain							**Avg.**	VisDA-2017: Acc. (%) on the target domain	
	Rel→Cli	Rel→Pnt	Rel→Skt	Pnt→Cli	Pnt→Rel	Cli→Skt	Skt→Pnt			
1-shot labeled source sample / 3-shots labeled source sample									**0.1% labeled**	**1% labeled**
Source Only	23.1/34.7	34.2/46.5	19.5/27.8	18.7/27.6	30.8/51.6	14.8/27.4	15.2/25.5	24.2/34.4	50.1	55.3
MME [48] (ICCV'19)	13.8/22.8	29.2/46.5	9.7/14.5	16.0/25.1	26.0/50.0	13.4/20.1	14.4/24.9	17.5/29.1	55.6	69.4
CDAN [30] (CVPR'18)	16.0/30.0	25.7/40.1	12.9/21.7	12.6/21.4	19.5/40.8	7.2/17.1	8.0/19.7	14.6/27.3	58.0	61.5
DALN [31] (CVPR'22)	25.6/34.7	29.3/46.5	17.6/26.6	15.9/28.3	22.1/47.4	14.7/22.5	13.2/23.8	19.8/32.8	61.7	62.1
DGWA [10] (TMM'24)	27.8/38.2	32.2/48/9	21.3/30.1	18.6/33.5	26.0/50.2	17.8/28.1	16.9/27.3	22.9/36.6	64.3	63.8
CDS [23] (arXiv'20)	16.7/35.0	24.4/43.8	11.1/36.7	14.1/34.1	15.9/36.8	13.4/31.1	19.0/34.5	16.4/36.0	34.2	67.5
PCS [24] (CVPR'21)	39.0/45.2	51.7/59.1	39.8/41.9	26.4/41.0	38.8/66.6	23.7/31.9	23.6/37.4	34.7/46.1	78.0	79.0
C-VisDiT [26] (ICCV'23)	39.5/45.7	48.8/57.4	38.7/45.7	26.7/42.5	43.3/68.2	26.5/33.6	27.5/43.8	35.9/48.1	79.2	80.5
ADCC	**41.8/47.6**	**51.8/59.5**	**40.2/47.9**	**29.1/43.9**	**45.6/70.4**	**29.7/35.8**	**29.4/46.3**	**38.2/50.2**	**80.5**	**81.7**

TABLE V
CONTRIBUTION OF EACH COMPONENT TO PERFORMANCE ON THE OFFICE31 DATASET.

$\mathcal{L}_{ICC}$	$\mathcal{L}_{\mathcal{I}}$	Office31: Acc. (%) on the target domain (1-shot/3-shots)						**Avg.**
		A→D	A→W	D→A	D→W	W→A	W→D	
✗	✗	63.8/78.6	71.4/83.5	77.0/77.8	90.3/94.3	72.9/76.3	91.5/96.2	77.8/84.5
✓	✗	74.3/84.0	70.6/87.6	77.6/78.5	92.9/95.2	77.5/79.8	94.2/97.4	81.2/87.1
✗	✓	73.7/82.9	70.0/87.3	77.2/77.5	92.5/95.0	77.3/78.2	94.1/97.1	80.8/86.3
✓	✓	**76.7/86.1**	**74.8/89.7**	**78.1/79.3**	**94.7/96.3**	**79.8/81.0**	**95.9/98.3**	**83.0/89.0**

[10], and MDDIA [49]. It is noteworthy that the experimental results for the FUDA approaches are reported from their original papers, while the results for the standard DA approaches are reported using their publicly available code.

B. Comparisons to the State-of-the-art (SOTA)

1) Office31 and Office-Home: As shown in Tables II and III, comparing the standard UDA methods MME [48], CDAN [30], and MDDIA [49], we observe that the few-shot source domain cannot directly improve the classification performance by the standard UDA methods, which verifies that the FUDA setting is indeed challenging. Note that the current SOTA approach is C-VisDiT [26], in both the Office31 and Office-Home datasets, our ADCC surpasses C-VisDiT in all the experimental tasks and obtains the highest average classification performance (**83.0%** (1-shot) and **89.0%** (3-shots) in Office31 and **64.1%** (3% labeled source samples) and **67.0%** (6% labeled source samples) in Office-Home). Different from the baseline PCS [24] and the SOTA method C-VisDiT, ADCC further considers the relation between the labeled source data and the unlabeled source samples by the ICC module. The more accurate pseudo-labels generated by the ICC module for the unlabeled source data further boost the model's performance in exploring discriminative information. In addition, the C-VisDiT also adopted the mix-up technique to augment the feature representation, but our ADCC still outperforms the C-VisDiT since the process of domain alignment was optimized by the adaptive dispersal strategy.

2) miniDomainNet and VisDA-2017: Table IV presents the classification accuracies on the miniDomainNet and VisDA-2017 datasets, respectively. It can be observed that our ADCC can achieve significant progress compare with the SOTA methods in both the miniDomainNet and VisDA-2017 datasets. Specifically, ADCC outperforms the SOTA methods in most of the experimental tasks and achieves the best classification accuracies (**38.2%** (1-shot) and **50.2%** (3-shots) in miniDomainNet and **80.5%** (0.1% labeled source samples) and **81.7%** (1% labeled source samples) in VisDA-2017). Overall, these results strongly demonstrate that intra-domain collaborative clustering and the cross-domain adaptive dispersal strategy are suitable for improving domain transfer with few-shot labeled source domain samples.

C. Further Analysis

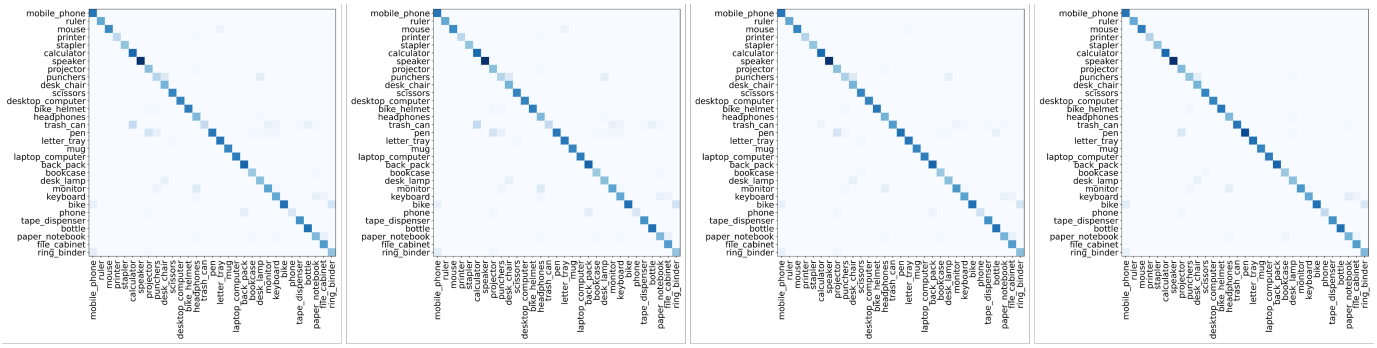
1) Ablation Study: To comprehensively assess the efficacy of each component in our proposed approach, we conducted a thorough series of ablation studies on the Office31 dataset. The pertinent results are presented in Table V. It is evident that all the proposed components contribute significantly to the overall performance improvement. Specifically, comparing the results of the second row and the third row, the ICC significantly boosts the classification performance since the ICC module generates more accurate pseudo-labels. It facilitates the model in exploring a wider range of discriminative information within the source domain. Although the performance enhancement achieved by only adopting the cross-domain adaptive dispersal strategy is not significant, the combination of ICC and the cross-domain adaptive dispersal strategy significantly improves the model performance. This is because the cross-domain adaptive dispersal strategy promotes the transfer of additional knowledge learned by the ICC module to the target domain. These obtained results further demonstrate that each

TABLE VI
COMPARISON OF DIFFERENT SAMPLING STRATEGY FOR MIX-UP RATIO ON THE OFFICE31 DATASET.

Settings	Office31: Acc. (%) on the target domain (1-shot/3-shots)						Avg.
	A→D	A→W	D→A	D→W	W→A	W→D	
None	74.3/84.0	70.6/87.6	77.6/78.5	92.9/95.2	77.5/79.8	94.2/97.4	81.2/87.1
Closer to the Source	74.9/85.1	73.3/88.1	76.9/79.0	93.8/95.9	79.3/80.6	94.7/97.9	82.2/87.8
Random Sampling	75.1/82.9	73.5/87.7	76.8/78.1	93.9/95.7	77.8/78.5	94.7/97.9	82.0/87.5
Linear Sampling	75.8/85.5	74.2/88.9	77.5/78.7	94.3/96.1	78.4/80.3	95.1/98.3	82.6/88.0
Adaptive Adjustment	**76.7/86.1**	**74.8/89.7**	**78.1/79.3**	**94.7/96.3**	**79.8/81.0**	**95.9/98.3**	**83.0/89.0**

TABLE VII
COMPARISON OF DIFFERENT LABELED SOURCE SAMPLES ON THE MINIDOMAINNET DATASET.

Method	1-shot	2-shots	3-shots	4-shots	5-shots	6-shots	7-shots	8-shots	9-shots	10-shots	fully-labeled
PCS	34.7	40.5	46.1	48.6	51.5	53.2	57.8	59.7	61.2	61.5	61.7
C-VisDiT	35.9	42.3	48.1	50.4	54.3	55.9	59.2	61.4	63.7	64.0	64.6
ADCC	**38.2**	**44.8**	**50.2**	**53.4**	**56.2**	**58.5**	**60.7**	**63.8**	**65.5**	**66.2**	**67.1**



(a) Source only (b) PCS (c) C-VisDiT (d) ADCC

Fig. 3. The confusion matrices of different methods on task A→W (Office31).

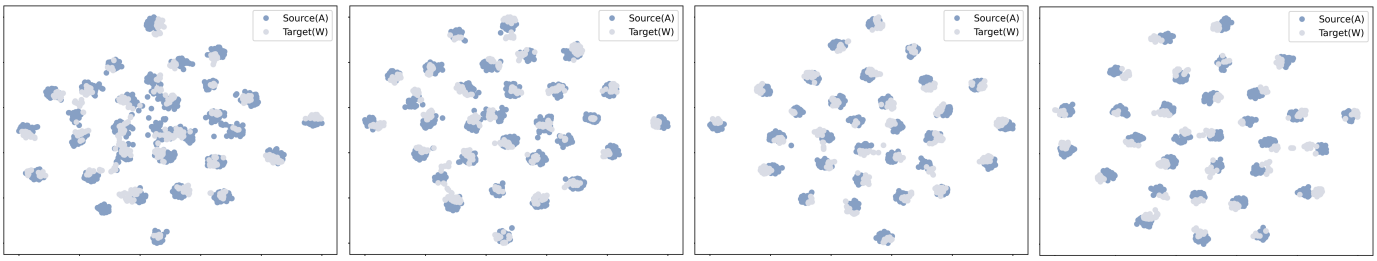
component of our proposed method is effective for FUDA issue.

2) Superiority of the Adaptive Dispersal Strategy: Since the cross-domain adaptive dispersal strategy plays a key role in cross-domain alignment, we further compare it with two alternative mix-up strategies: the general mix-up strategy [26], where the mix-up ratio is sampled from a beta distribution $Beta(0, 1)$, and the linear sampling mix-up ratio strategy, which uses a linear sampling approach as a substitute for the adaptive mix-up strategy. The experiments were conducted on the Office31 dataset, and the results are presented in Table VI. Notably, we observe that all three sampling strategies improve the model's performance compared to the baseline model. But the performance of our adaptive adjustment strategy is significantly superior to that of the random sampling and linear sampling strategies since the adaptive adjustment strategy can better align the source and target distributions. In addition, we also evaluate the initially setting of the intermediate domain. From Table VI, we can observe that the performance of set intermediate domain closer to the target domain (the last row of Table VI) is better than that of set intermediate domain closer to the source domain (the fourth row of Table VI).

3) Comparisons on Varying Numbers of Source Labels: To further evaluate the effectiveness of ADCC under varying numbers of labeled source samples, we conducted experiments as shown in Table VII. Specifically, the scale of labeled source samples ranges from 1-shot to 10-shots, as well as a fully-labeled source scenario for comparison. From the experimental results, we observe that as the number of labeled source samples increases, the classification accuracies of PCS, C-VisDiT, and ADCC also improve. Moreover, across 1-10 shot scenarios and even in the fully-labeled case, ADCC consistently achieves superior performance. This can be attributed to the ICC module effectively leveraging the information from labeled samples and the adaptive dispersal strategy ensuring smoother domain alignment.

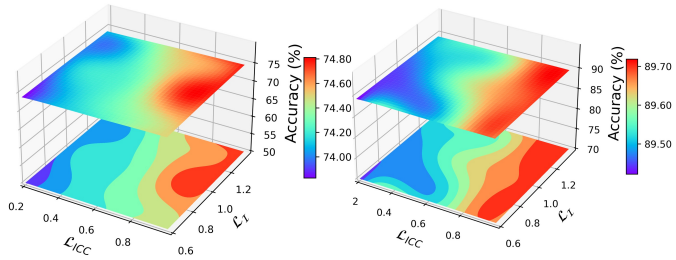
4) Confusion Matrix: Figure 3 depicts the confusion matrices for the classifier learned by Source-only, PCS, C-VisDiT, and ADCC on the Office31 dataset. The Source-only model yields numerous incorrect predictions owing to the domain shift between the source and target domains. After adaptation by ADCC, most of the wrong predictions were corrected. Compared to Source-only, our method generates more correct predictions that are located on the main diagonal of the confusion matrix, demonstrating that our method learns more discriminative information for the target domain.

5) Feature Visualization: Figure 4 illustrates the t-SNE [57] feature visualization for the embeddings generated by



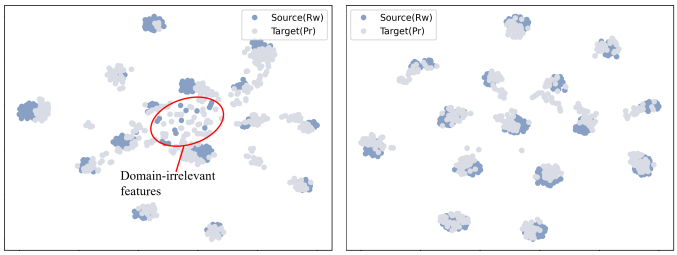
(a) Source only (b) PCS (c) C-VisDiT (d) ADCC

Fig. 4. The t-SNE visualization of the features on the Office31 dataset.



(a) A→W (1-shot) (b) A→W (3-shots)

Fig. 5. The analysis of the ADCC parameters on task A→W (Office31).



(a) Before filtering out domain-irrelevant features (b) After filtering out domain-irrelevant features

Fig. 6. t-SNE visualization of feature distributions on the Office-Home dataset. (a) Before filtering out domain-irrelevant features, the source and target feature distributions overlap significantly, indicating the presence of substantial domain-irrelevant information. (b) After filtering out domain-irrelevant features using our ADCC framework, the source and target feature distributions become more compact and well-aligned, effectively mitigating the impact of domain-irrelevant information.

the Source-only, PCS, C-VisDiT, and ADCC model on the Office31 dataset. We can observe that the distribution of the target domain cannot be effectively clustered by the source-only model. After adaptation by ADCC, features belonging to the same class exhibit improved clustering, while features from distinct classes are better separated. The obtained results of feature visualization further demonstrate the effectiveness of our proposed method.

6) Hyper-parameter Sensitivity: We further conduct the hyper-parameter analysis on λ_{ICC} and $\lambda_{\mathcal{I}}$ in Eq. (20) on the Office31 dataset. We select different values of the parameters i.e., $\lambda_{ICC} = \{0.2, 0.4, 0.6, 0.8, 1.0\}$ and $\lambda_{\mathcal{I}} = \{0.6, 0.8, 1.0, 1.2, 1.4\}$. As Figures 5 (a) and (b) shown, the model is slightly sensitive to λ_{ICC}, and the best choice of λ_{ICC} is 0.6. Meanwhile, the model is not sensitive to $\lambda_{\mathcal{I}}$, and can attain competitive outcomes across a broad spectrum of hyper-parameter values.

7) Additional Visualization Analysis for Domain-irrelevant Features: We also conduct visualization experiments on the Office-Home dataset to illustrate the negative effects of domain-irrelevant features. For clearer visualization, we selected 15 representative categories from the dataset for analysis. Specifically, we compare the feature distributions of the source and target domains before and after filtering out domain-irrelevant information. The results are presented in Figure 6. As shown in Figure 6(a), when samples with high domain-irrelevant content are not filtered, there is substantial overlap between the source and target feature distributions. This overlap suggests the presence of significant domain-irrelevant information, which hinders effective alignment and may lead to negative transfer. In contrast, Figure 6(b) illustrates that our ADCC filtering out domain-irrelevant features, yields more compact and well-aligned feature distributions, thereby effectively mitigating the impact of domain-irrelevant information.

V. Conclusion

In this paper, we focus on the more challenge few-shot unsupervised domain adaptation (FUDA) issue in which only a few samples in the source domain are annotated while the remaining source and target domains samples are unlabeled. Considering the inadequacy of labeled samples in the source domain, we propose the intra-domain collaborative clustering to explore more discriminative information by bridging the labeled source samples and the unlabeled source samples. Furthermore, the cross-domain adaptive dispersal strategy effectively alleviates the impact of domain-irrelevant information between the source and target domains. Extensive experiments demonstrate that our method achieves superior performance on four few-shot unsupervised domain adaptation benchmarks.

References

[1] Z. Li, H. Tang, Z. Peng, G.-J. Qi, and J. Tang, "Knowledge-guided semantic transfer network for few-shot image recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, 2024. doi: 10.1109/TNNLS.2023.3240195.

[2] Y. Zhang, W. Li, W. Sun, R. Tao, and Q. Du, "Single-source domain expansion network for cross-scene hyperspectral image classification," *IEEE Trans. Image Process*, vol. 32, pp. 1498–1512, 2023.

[3] M. Chen, Z. Zheng, Y. Yang, and T.-S. Chua, "Pipa: Pixel- and patch-wise self-supervised learning for domain adaptative semantic segmentation," in *Proc. ACM Int. Conf. Multimedia*, pp. 1905–1914, 2023.

[4] Q. Nguyen, T. Vu, A. Tran, and K. Nguyen, "Dataset diffusion: Diffusion-based synthetic data generation for pixel-level semantic segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 76872–76892, 2024.

[5] Y. Pu, W. Liang, Y. Hao, Y. YUAN, Y. Yang, C. Zhang, H. Hu, and G. Huang, "Rank-detr for high quality object detection," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 16100–16113, 2024.

[6] Z. Wang, Y. Li, X. Chen, S. Lim, A. Torralba, H. Zhao, and S. Wang, "Detecting everything in the open world: Towards universal object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 11433–11443, 2023.

[7] S. Bendavid, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. W. Vaughan, "A theory of learning from different domains," *Mach. Learn*, vol. 79, no. 1-2, pp. 151–175, 2010.

[8] S. Ben-David, J. Blitzer, K. Crammer, and F. Pereira, "Analysis of representations for domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 137–144, 2007.

[9] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, 2010.

[10] Y. Lu, H. Huang, B. Zeng, Z. Lai, and X. Li, "Multi-source and multi-target domain adaptation based on dynamic generator with attention," *IEEE Trans. Multimedia*, vol. 26, pp. 6891–6905, 2024.

[11] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 2962–2971, 2017.

[12] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Unsupervised domain adaptation with residual transfer networks," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 136–144, 2016.

[13] D. Peng, Q. Ke, A. Ambikapathi, Y. Yazici, Y. Lei, and J. Liu, "Unsupervised domain adaptation via domain-adaptive diffusion," *IEEE Trans. Image Process*, 2024. doi: 10.1109/TIP.2024.3424985.

[14] Z. Du and J. Li, "Diffusion-based probabilistic uncertainty estimation for active domain adaptation," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 17129–17155, 2024.

[15] Z. Deng, K. Zhou, D. Li, J. He, Y.-Z. Song, and T. Xiang, "Dynamic instance domain adaptation," *IEEE Trans. Image Process*, vol. 31, pp. 4585–4597, 2022.

[16] Y. Li, L. Yuan, Y. Chen, P. Wang, and N. Vasconcelos, "Dynamic transfer for multi-source domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 10998–11007, 2021.

[17] C.-X. Ren, Y.-W. Luo, and D. Dai, "Buresnet: Conditional bures metric for transferable representation learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 4, pp. 4198–4213, 2022.

[18] Z. Wang, L. Shen, M. Xu, M. Yu, K. Wang, and Y. Lin, "Domain adaptation for underwater image enhancement," *IEEE Trans. Image Process*, vol. 32, pp. 1442–1457, 2023.

[19] D. Wang, Y. Zhang, K. Zhang, and L. Wang, "Focalmix: Semi-supervised learning for 3d medical image detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 3950–3959, 2020.

[20] V. I. Butoi, J. J. G. Ortiz, T. Ma, M. R. Sabuncu, and J. Guttag, "Universeg: Universal medical image segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 21438–21451, 2023.

[21] T. Zhang, J. Chen, F. Li, T. Pan, and S. He, "A small sample focused intelligent fault diagnosis scheme of machines via multimodules learning with gradient penalized generative adversarial networks," *IEEE Trans. Ind. Electron*, vol. 68, no. 10, pp. 10130–10141, 2020.

[22] Y. Dong, Y. Li, H. Zheng, R. Wang, and M. Xu, "A new dynamic model and transfer learning based intelligent fault diagnosis framework for rolling element bearings race faults: Solving the small sample pproblem," *ISA Trans.*, vol. 121, pp. 327–348, 2022.

[23] D. Kim, K. Saito, T.-H. Oh, B. A. Plummer, S. Sclaroff, and K. Saenko, "Cross-domain self-supervised learning for domain adaptation with few source labels," *arXiv preprint*, 2020. arXiv:2003.08264.

[24] X. Yue, Z. Zheng, S. Zhang, and Y. Guo, "Prototypical cross-domain self-supervised learning for few-shot unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 13834–13844, 2021.

[25] S. Huang, W. Yang, L. Wang, L. Zhou, and M. Yang, "Few-shot unsupervised domain adaptation with image-to-class sparse similarity encoding," in *Proc. ACM Int. Conf. Multimedia*, pp. 677–685, 2021.

[26] Y. Xiong, H. Chen, Z. Lin, S. Zhao, and G. Ding, "Confidence-based visual dispersal for few-shot unsupervised domain adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 11621–11631, 2023.

[27] B. Xie, S. Li, F. Lv, C. H. Liu, G. Wang, and D. Wu, "A collaborative alignment framework of transferable knowledge extraction for unsupervised domain adaptation," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 7, pp. 6518–6533, 2023.

[28] M. Long, Y. Cao, Z. Cao, J. Wang, and M. I. Jordan, "Transferable representation learning with deep adaptation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 12, pp. 3071–3085, 2019.

[29] M. Long, H. Zhu, J. Wang, and M. I. Jordan, "Deep transfer learning with joint adaptation networks," in *Proc. Int. Conf. Mach. Learn.*, pp. 2208–2217, 2017.

[30] M. Long, Z. Cao, J. Wang, and M. I. Jordan, "Conditional adversarial domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 1640–1650, 2018.

[31] L. Chen, H. Chen, Z. Wei, X. Jin, X. Tan, Y. Jin, and E. Chen, "Reusing the task-specific classifier as a discriminator: Discriminator-free adversarial domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 7171–7181, 2022.

[32] S. Li, M. Xie, F. Lv, C. H. Liu, J. Liang, C. Qin, and W. Li, "Semantic concentration for domain adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 9102–9112, 2021.

[33] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, "Domain-adversarial training of neural networks," *J. Mach. Learn Res.*, vol. 17, no. 59, pp. 1–35, 2016.

[34] L. Yu, W. Yang, S. Huang, L. Wang, and M. Yang, "High-level semantic feature matters few-shot unsupervised domain adaptation," in *Proc. AAAI Conf. Artif. Intell.*, pp. 11025–11033, 2023.

[35] J. Zhang, H. Chao, A. Dhurandhar, P.-Y. Chen, A. Tajer, Y. Xu, and P. Yan, "Spectral adversarial mixup for few-shot unsupervised domain adaptation," in *Proc. Int. Conf. Med. Image Comput. Comput. Assisted Intervention*, pp. 728–738, 2023.

[36] L. Gong, W. Zhang, M. Li, J. Zhang, and Z. Zhang, "Prototype-augmented contrastive learning for few-shot unsupervised domain adaptation," in *Proc. Int. Conf. Konwl. Sci. Eng. Manag.*, pp. 197–210, 2023.

[37] J. Ling, L. Liao, M. Yang, and J. Shuai, "Semi-supervised few-shot learning via multi-factor clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 14564–14574, 2021.

[38] H.-J. Ye, L. Ming, D.-C. Zhan, and W.-L. Chao, "Few-shot learning with a strong teacher," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 46, no. 3, pp. 1425–1440, 2024.

[39] R. Subramanyam, M. Heimann, J. S. Thathachar, R. Anirudh, and J. Thiagarajan, "Contrastive knowledge-augmented meta-learning for few-shot classification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, pp. 2479–2487, 2023.

[40] C. Simon, P. Koniusz, and M. Harandi, "Meta-learning for multi-label few-shot classification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, pp. 3951–3960, 2022.

[41] Z. Ji, Z. Hou, X. Liu, Y. Pang, and X. Li, "Memorizing complementation network for few-shot class-incremental learning," *IEEE Trans. Image Process*, vol. 32, pp. 937–948, 2023.

[42] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 471–478, 2011.

[43] Y. Wu, D. Inkpen, and A. El-Roby, "Dual mixup regularized learning for adversarial domain adaptation," in *Proc. Eur. Conf. Comput. Vis.*, pp. 540–555, 2020.

[44] M. Xu, J. Zhang, B. Ni, T. Li, C. Wang, Q. Tian, and W. Zhang, "Adversarial domain adaptation with domain mixup," in *Proc. AAAI Conf. Artif. Intell.*, pp. 6502–6509, 2020.

[45] Y. Shu, Z. Cao, C. Wang, J. Wang, and M. Long, "Open domain generalization with domain-augmented meta-learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 9624–9633, 2021.

[46] Z. Mai, G. Hu, D. Chen, F. Shen, and H. T. Shen, "Metamixup: Learning adaptive interpolation policy of mixup with metalearning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 7, pp. 3050–3064, 2020.

[47] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proc. Int. Conf. Mach. Learn.*, pp. 41–48, 2009.

[48] K. Saito, D. Kim, S. Sclaroff, T. Darrell, and K. Saenko, "Semi-supervised domain adaptation via minimax entropy," in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 8050–8058, 2019.

[49] X. Jiang, Q. Lao, S. Matwin, and M. Havaei, "Implicit class-conditioned domain alignment for unsupervised domain adaptation," in *Proc. Int. Conf. Mach. Learn.*, pp. 4816–4827, 2020.

[50] K. Saenko, B. Kulis, M. Fritz, and T. Darrell, "Adapting visual category models to new domains," in *Proc. Eur. Conf. Comput. Vis.*, pp. 213–226, 2010.

[51] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan, "Deep hashing network for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 5018–5027, 2017.

[52] Y. Y. Kaiyang Zhou, Y. Qiao, and T. Xiang, "Domain adaptive ensemble learning," *IEEE Trans. Image Process*, vol. 30, pp. 8008–8018, 2021.

[53] X. Peng, Q. Bai, X. Xia, Z. Huang, K. Saenko, and B. Wang, "Moment matching for multi-source domain adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 1406–1415, 2019.
[54] X. Peng, B. Usman, N. Kaushik, J. Hoffman, D. Wang, and K. Saenko, "Visda: The visual domain adaptation challenge," *arXiv preprint*, 2017. arXiv:1710.06924.
[55] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, and L. Antiga, "Pytorch: An imperative style, high-performance deep learning library," in *Proc. Adv. Neural Inf. Process. Syst.*, pp. 8026–8037, 2019.
[56] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pp. 770–778, 2016.
[57] L. V. D. Maaten and G. Hinton, "Visualizing data using t-sne," *J. Mach. Learn Res.*, vol. 9, pp. 2579–2605, 2008.

Yuwu Lu (Senior member, IEEE) received the Ph.D. degree in computer science and technology from the Harbin Institute of Technology, Harbin, China, in 2015. He is a Postdoctoral Fellow with Tsinghua Shenzhen International Graduate School, Shenzhen, China. He is an Associate Professor with South China Normal University, Guangzhou, China. He is an Associate Editor for *Pattern Recognition*. He has authored or coauthored more than 40 scientific papers, including IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON MULTIMEDIA, *Pattern Recognition*. His current research interests include pattern recognition and machine learning.

Haoyu Huang (Student member, IEEE) received the B.S. degree in software engineering from Jilin University Zhuhai College, Zhuhai, China, in 2022. He is currently pursuing the M.S. degree in software engineering from South China Normal University, Foshan, China. His current research interests include deep learning and transfer learning.

Wai Keung Wong received the Ph.D. degree from Hong Kong Polytechnic University, Hong Kong, SAR, in 2002. He is a Cheng Yik Hung Endowed Professor in Fashion and currently serving as the CEO and the Centre Director of the Laboratory for Artificial Intelligence in Design. He has published over 150 scientific articles in refereed journals, including the IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Image Processing, IEEE Transactions on Cybernetics, and Pattern Recognition. His recent research interests include pattern recognition, feature extraction, machine learning, and defect detection.

Xue Hu received the B.S. degree in software engineering from Hubei University of Economics, Wuhan, China, in 2022. She is currently pursuing the M.S. degree in software engineering from South China Normal University, Foshan, China. Her current research interests include deep learning and transfer learning.

Zhihui Lai (Member, IEEE) received the B.S. degree in mathematics from South China Normal University, M.S. degree from Jinan University, and the Ph.D. degree in pattern recognition and intelligence system from Nanjing University of Science and Technology (NUST), China, in 2002, 2007 and 2011, respectively. He has been a Research Associate, Postdoctoral Fellow and Research Fellow at The Hong Kong Polytechnic University. His research interests include face recognition, image processing and content-based image retrieval, pattern recognition, compressive sense, human vision modelization and applications in the fields of intelligent robot research.

Xuelong Li (Fellow, IEEE) is with the Institute of Artiffcial Intelligence (TeleAI), China Telecom Corp Ltd, 31 Jinrong Street, Beijing 100033, P. R. China.