

# Federated Source-free Domain Adaptation for Classification: Weighted Cluster Aggregation for Unlabeled Data

Junki Mori  
NEC Corporation, Kyoto University  
Kawasaki, Japan  
junki.mori@nec.com

Kosuke Kihara  
NEC Corporation  
Kawasaki, Japan  
kosuke-kihara@nec.com

Taiki Miyagawa  
NEC Corporation  
Kawasaki, Japan  
miyagawataik@nec.com

Akinori F. Ebihara  
NEC Corporation  
Kawasaki, Japan  
aebihara@nec.com

Isamu Teranishi  
NEC Corporation  
Kawasaki, Japan  
teranisi@nec.com

Hisashi Kashima  
Kyoto University  
Kyoto, Japan  
kashima@i.kyoto-u.ac.jp

## Abstract

Federated learning (FL) commonly assumes that the server or some clients have labeled data, which is often impractical due to annotation costs and privacy concerns. Addressing this problem, we focus on a source-free domain adaptation task, where (1) the server holds a pre-trained model on labeled source domain data, (2) clients possess only unlabeled data from various target domains, and (3) the server and clients cannot access the source data in the adaptation phase. This task is known as *Federated source-free Domain Adaptation (FFREEDA)*. Specifically, we focus on classification tasks, while the previous work solely studies semantic segmentation. Our contribution is the novel *Federated learning with Weighted Cluster Aggregation (FedWCA)* method, designed to mitigate both domain shifts and privacy concerns with only unlabeled data. FedWCA comprises three phases: private and parameter-free clustering of clients to obtain domain-specific global models on the server, weighted aggregation of the global models for the clustered clients, and local domain adaptation with pseudo-labeling. Experimental results show that FedWCA surpasses several existing methods and baselines in FFREEDA, establishing its effectiveness and practicality.

## 1. Introduction

Most existing studies on federated learning (FL), a privacy-preserving machine learning paradigm for distributed data, assume that labeled data is available at each client. However, this may not be the case for vision tasks which often necessitate cumbersome manual data annotation requiring domain expertise. For instance, in applying

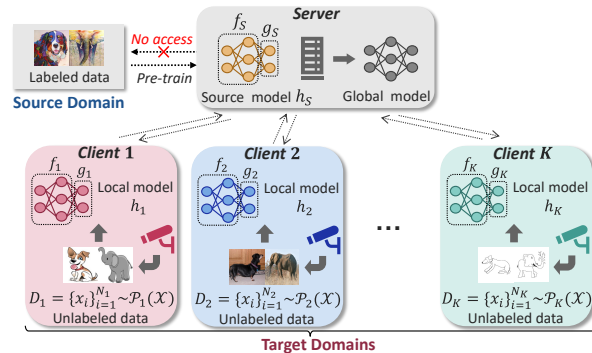


Figure 1. **Overview of our problem setting: FFREEDA [45].** A pre-trained model on a labeled *source* domain dataset exists on the server, which becomes inaccessible following pre-training (*source-free*). Clients possess unlabeled datasets from specific *target* domains, which may or may not be identical. We aim to train personalized models tailored to each client’s domain.

FL to multi-location surveillance cameras, it is burdensome for administrators to manually annotate each image. While a server, typically hosted by organizations, is more likely to access labeled data, the included biometric information such as faces, makes direct use of this data in FL problematic due to privacy concerns. The situation also exemplifies the challenges of differing domain characteristics, like brightness and background, based on the camera’s location.

Building on the preceding example, this study explores the practical yet less-investigated FL setting of *Federated source-Free Domain Adaptation (FFREEDA)* [45], specifically for classification tasks (Fig. 1). In FFREEDA, a server possesses a model pre-trained on labeled data from a source domain, but this data is not usable in later FL stages due to

privacy issues or immediate deletion after pre-training. In addition, each client holds limited unlabeled data in a target domain, which may greatly differ from the source domain and the domains of other clients, thus creating an unsupervised multi-target setting. FFREEDA aims for each client to train a classification model tailored to its domain, leveraging the source model and the knowledge from other clients.

We tackle two main challenges of FFREEDA: (1) fully unsupervised FL after pre-training and (2) domain shifts among clients. We propose a novel FFREEDA method for classification, *Federated Learning with Weighted Cluster Aggregation* (FedWCA). This overcomes issues in the prior FFREEDA method [45] for semantic segmentation, including privacy issues connected to image style feature sharing during client clustering and limited performance (Sec. 5) due to a cluster-restricted model aggregation and a simple pseudo-labeling. Specifically, FedWCA comprises three key components. Firstly, to avert performance degradation from domain shift (2), we cluster clients based on the initial layer parameters of each client’s locally adapted feature extractor, which contains domain information. This process relies only on model parameters, eliminating the need for additional information sharing and hyperparameters. To further boost performance, we introduce the Weighted Cluster Aggregation (WCA) module, enabling clients to combine models from each cluster (domain) using personal weights. A novel metric is defined for computing cluster weights based solely on unlabeled data, considering the benefits for both individual clients and the overall group. Finally, handling unlabeled target data (1) is achieved through prototype-based pseudo-labeling [29]. We improve it to retain global knowledge and produce more reliable pseudo-labels. Empirical validation confirms the efficacy of these components, with the complete method outperforming all baselines, as detailed in Sec. 5.

This work primarily contributes as follows:

- We introduce a novel FFREEDA method for classification, named FedWCA, with three main components: (1) a private, parameter-free client clustering module, (2) a weighted cluster aggregation module for domain knowledge integration, and (3) a local adaptation module using an improved pseudo-labeling strategy.
- Empirical results show that FedWCA significantly enhances performance over all baselines, including the prior method [45], across three multi-domain benchmark datasets: Digit-Five, PACS, and Office-Home.

## 2. Related Work

### 2.1. Source-free Domain Adaptation

This paper extends *Source-free domain adaptation* (SFDA), a type of unsupervised domain adaptation ap-

proach, to the FL setting. SFDA involves adapting a pre-trained source model to an unlabeled target domain without accessing the source data. A fundamental SFDA method is pseudo-labeling for unlabeled target data [6, 24, 29, 31, 40, 54]. A prototype-based pseudo-labeling method introduced by *Source Hypothesis Transfer* (SHOT) [29], is commonly employed to derive pseudo-labels, minimizing label noise induced by domain shifts [6, 40, 54].

### 2.2. Federated Domain Adaptation

FADA [38] is the pioneering work addressing *Federated Domain Adaptation* (FDA) for a single target domain. Subsequent studies [9, 16, 20, 48] have also focused on a multi-source-single-target setting. DualAdapt [58] explores a single-source-multi-target scenario, similar to this paper, where each client gets one target domain and the server holds a source domain. FRuDA [10] considers sequentially adding new target domains to the federated system. Unlike our source-free setting, these studies typically assume access to source data during FL (Tab. 1).

### 2.3. Federated Source-free Domain Adaptation

Our model is inspired by *Learning Across Domains and Devices* (LADD) [45], the sole study addressing FFREEDA, which focuses on semantic segmentation. In contrast, our model is designed for classification tasks, decreases the number of hyperparameters, alleviates privacy concerns by eliminating the need to transmit image style features from clients to servers for client clustering (Sec. 4.1), integrates all domain information with client-specific weights (Sec. 4.2), generates more reliable pseudo-labels (Sec. 4.3), and thus improves accuracy (Sec. 5).

### 2.4. Federated Learning with Non-i.i.d. Data

A major challenge in FL is data heterogeneity, namely the non-i.i.d. (independent and identically distributed) data distribution problem [27]. *Personalized FL* [49] addresses this by training models tailored to each client’s unique data distribution [4, 7, 17, 33, 35, 47, 61]. *Clustered FL*, a variant of personalized FL, also tackles non-i.i.d. data by clustering clients based on data distributions and aggregating models within clusters. However, existing clustered FL methods often require setting hyperparameters such as cluster numbers [2, 12, 22, 37, 43], thresholds [1, 59], iteration counts [52], and neighbor candidate list sizes [28]. Although these studies focus on non-i.i.d. data, they largely assume labeled data and emphasize label shifts more than domain shifts (Tab. 1).

### 2.5. Federated Learning with Unlabeled Data

In the realm of FL with unlabeled data, such as *semi-supervised FL*, two main settings exist: one where the server has labeled data and clients only have unlabeled data [5, 19, 26, 64] and another where some clients possess

Table 1. **Related problem settings.** Our study focuses on classification tasks, tackling three distinct problem types: "No Labels," "Domain Shift," and "Personalization."  $\Delta$  denotes specific characteristics depending on FL paradigms: while most Personalized FL research concentrates on label shifts, certain studies like [47, 50] also tackle domain shifts. Semi-Supervised FL literature typically involves unlabeled data, presuming the presence of labeled data either on the server or within a subset of clients, a scenario also observed in some FDA studies.

FL paradigm	No labels	Domain shift	Personalization	Task
Personalized FL [4, 7, 17, 33, 35, 47, 50, 61]		$\Delta$	$\checkmark$	Classification
Semi-supervised FL [5, 8, 19, 26, 30, 64]	$\Delta$			Classification
Unsupervised FL [3, 21, 32, 44, 60]	$\checkmark$			Representation learning/Clustering
FDA [9, 16, 20, 38, 48]	$\Delta$	$\checkmark$		Classification
[10, 58]	$\Delta$	$\checkmark$	$\checkmark$	Classification
FFREEDA LADD [45]	$\checkmark$	$\checkmark$	$\checkmark$	Semantic segmentation
<b>Ours</b>	$\checkmark$	$\checkmark$	$\checkmark$	Classification

fully labeled data while others have entirely unlabeled data [8, 30, 55]. Our work explores a fully unsupervised setting for both server and client sides after the source pre-training phase. *Unsupervised FL*, which treats fully unsupervised settings throughout training, typically focuses on representation learning [21, 32, 60] or clustering tasks [3, 44], unlike our emphasis on classification tasks (Tab. 1).

### 3. Problem Formulation

Our FFREEDA is illustrated in Fig. 1. In our FFREEDA, a server and  $K$  clients (indexed by  $k$ ) address an  $M$ -way classification problem with a feature space  $\mathcal{X}$  and a label space  $\mathcal{Y}$ . Each client  $k$  holds an unlabeled dataset  $D_k = \{x_i\}_{i=1}^{N_k} \sim \mathcal{P}_k(\mathcal{X})$  from multiple target domains (client side in Fig. 1), where  $\mathcal{P}_k$  represent the respective data distributions. Distinct domain shifts exist among the clients, denoted by  $\mathcal{P}_k \neq \mathcal{P}_l$ . Note that certain clients may share the same domain.

The server has a source model  $h_S: \mathcal{X} \rightarrow \mathbb{R}^M$  pre-trained on a labeled dataset from the source domain whose data distribution differs from  $\mathcal{P}_k$  for any  $k$  (server side in Fig. 1). The labeled source dataset is discarded or retained privately, preventing further access. The model  $h_S$  has two modules:  $h_S = g_S \circ f_S$ , where  $f_S: \mathcal{X} \rightarrow \mathbb{R}^q$  is a feature extractor,  $g_S: \mathbb{R}^q \rightarrow \mathbb{R}^M$  is a classifier, and  $\circ$  denotes the composition operation. The feature extractor  $f_S$  converts a sample  $x$  into a  $q$ -dimensional feature vector  $f_S(x)$ , while the classifier  $g_S$  assigns class scores to  $f_S(x)$ , with  $(g_S(f_S(x)))_m$  denoting the score for class  $m$ . The final linear layer of a neural network, typically viewed as the classifier  $g_S$ , has the weight matrix  $W = [w_1, \dots, w_M] \in \mathbb{R}^{q \times M}$ , where  $w_m$  is the classifier vector for class  $m$ .

The aim of FFREEDA is to evolve from the source model  $h_S$  to learn multiple personalized classification models  $h_k = g_k \circ f_k$ , each tailored to the respective client distributions  $\mathcal{P}_k(\mathcal{X})$ , while encouraging collaboration among clients. In each communication round  $r = 0, \dots, R - 1$ , clients train their models  $h_k$  locally for  $E$  epochs and send them to the server for aggregation to obtain global models.

See Tab. 7 in App. A for a full list of notation.

## 4. Method

FedWCA comprises three main components: (1) private and parameter-free client clustering, (2) weighted cluster aggregation, and (3) local adaptation via pseudo-labeling. In line with prevalent SFDA practices [6, 29, 40], we fix the classifier during FL, anticipating that the source model’s classifier has developed a domain-independent capability for the task. See App. B.1 for the pseudo-code of FedWCA.

### 4.1. Private and Parameter-free Client Clustering

To manage domain shifts among clients, our method clusters clients privately and without parameters as a first step, based on locally learned feature extractor parameters  $f_k$  by each client. This approach removes the need for server access to raw client data and requires only a one-time initial clustering operation.

Unlike traditional clustered FL methods that require hyperparameters such as the number of clusters, our strategy employs a parameter-free clustering: *First Integer Neighbor indices producing a Clustering Hierarchy* (FINCH) [42]. We use cosine similarity between clients’ model parameters to identify the nearest neighbor  $\kappa_k$  for each client  $k$ . The adjacency matrix is defined as  $A(k, l) = 1$  if  $l = \kappa_k$ ,  $k = \kappa_l$ ,  $\kappa_k = \kappa_l$ , and  $A(k, l) = 0$  otherwise. Clients linked by this matrix are grouped into the same cluster.

Given the typically large number of parameters in a feature extractor, using all parameters for distance evaluation can lead to a single cluster due to the curse of dimensionality (see Sec. 5.2.2). To mitigate this, we focus on the shallow layers where domain knowledge is stored [14]. Specifically, we only use the first-layer parameters as this layer, the furthest from the common classifier among clients, best represents domain-specific information (Sec. 5.2.2). Consequently, the server can derive  $C$  clusters (indexed by  $c$ ) and generate cluster models  $f_c$  by averaging the local models in each cluster. In the next section, we construct each client’s initial model for the next round using the cluster models.

## 4.2. Weighted Cluster Aggregation

The core of our FedWCA method is the weighted cluster aggregation, depicted in Fig. 2. In conventional clustered FL methods, the server distributes the cluster model  $f_{c_k}$  to client  $k$ , where  $c_k$  is the cluster index assigned to client  $k$ , for use as the initial model for the next round. This approach limits clients to knowledge solely from their assigned cluster, challenging clients in clusters with few data points or whose domains are significantly different from the source domain. To overcome this, our strategy enables clients to blend all cluster models  $f_c$  using specific cluster weights  $\mathbf{v}_k = (v_{k,1}, \dots, v_{k,C})$  (summing to 1), to form an initial model  $f_k^{\text{init}} = \sum_c v_{k,c} f_c$  in each round  $r^1$ . This blended model not only leverages cross-cluster knowledge but also corrects possible cluster assignment errors. At the end of the training, each client uses its cluster’s model for the performance evaluation. The specific cluster weights calculation method is described below.

**Local calculation of cluster weights.** Firstly, we explore a method for clients to compute cluster weights locally, vital for model personalization. The underlying principle is to assign higher weights to those cluster models that are most beneficial for a client’s domain.

Considering that an effective feature extractor converges samples towards a corresponding class-specific classifier vector defined in Sec. 3, which effectively acts as a prototype in the feature space, we introduce new metrics  $\alpha_k = (\alpha_{k,1}, \dots, \alpha_{k,C})$  to represent the cluster weights  $\mathbf{v}_k$  for each client  $k$ . First, client  $k$  evaluates the cosine similarity  $\cos(f_c(x_i), w_{m(i)})$  between the feature extractor output  $f_c(x_i)$  and the corresponding class’ classifier vector  $w_{m(i)}$ , where  $m(i) = \arg \max_m \cos(f_c(x_i), w_m)$  represents the class whose classifier vector is closest to  $f_c(x_i)$ . The similarity scores across all samples are averaged to compute  $I_{k,c} = \sum_{i=1}^{N_k} \cos(f_c(x_i), w_{m(i)}) / N_k$  for each cluster  $c$  and  $\alpha_k$  are derived using a temperature parameter  $T_a$  as follows:

$$(\alpha_{k,1}, \dots, \alpha_{k,C}) = \text{softmax}(I_{k,1}/T_a, \dots, I_{k,C}/T_a), \quad (1)$$

where ‘softmax’ is the softmax function.

However, this metric focuses solely on the benefits to the individual client, without considering the overall advantages across the clients. Such a client-centric approach may not necessarily contribute positively to the other clients.

**Global and local calculation of cluster weights.** We calculate the cluster weights, focusing on both individual client benefits and collective advantages. This involves using both cluster models  $f_c$  and ‘soft cluster models’  $\tilde{f}_c$ , which integrate benefits from other clusters.

For creating soft cluster models, we define a benefit metric  $A_{c' \rightarrow c}$ , with  $\sum_{c'} A_{c' \rightarrow c} = 1$ , indicating the rela-

<sup>1</sup>In the initial round,  $f_k^{\text{init}}$  aligns with the source feature extractor  $f_S$ , and  $g_k^{\text{init}}$  remains fixed to the source classifier  $g_S$  across all rounds.

tive advantage of cluster  $c$  for clients in cluster  $c'$ . This metric formulates new cluster models by weighting  $f_{c'}$  as  $\sum_{c'} A_{c' \rightarrow c} f_{c'}$ . To prevent loss of knowledge from the original cluster models, we introduce coefficients  $B_{c,0}$  and  $B_{c,1}$ , summing to 1, for each cluster. These coefficients balance the emphasis between the original cluster model  $f_c$  and the newly combined model  $\sum_{c'} A_{c' \rightarrow c} f_{c'}$ . The soft cluster models are thus given by

$$\tilde{f}_c = B_{c,0} f_c + B_{c,1} \sum_{c'} A_{c' \rightarrow c} f_{c'}. \quad (2)$$

$A_{c' \rightarrow c}$  and  $B_{i,c}$  are estimated later in this subsection.

Upon receiving the original cluster model  $f_{c_k}$  and all soft cluster models  $\tilde{f}_c$  from the server, each client  $k$  locally calculates cluster weights  $\alpha_k$  for the soft cluster models, leading to a composite model  $\bar{f}_k = \sum_c \alpha_{k,c} \tilde{f}_c$ . To adequately incorporate the original cluster model  $f_{c_k}$ , client  $k$  employs pseudo-performance weights  $\beta_k = (\beta_{k,0}, \beta_{k,1})$  (summing to 1) to balance between  $f_{c_k}$  and  $\bar{f}_k$ . These weights are derived using Soft Neighborhood Density (SND) [41], a method that assesses the overall model efficacy on unlabeled data by analyzing the entropy of similarity distributions between outputs:

$$(\beta_{k,0}, \beta_{k,1}) = \text{softmax}(S(h_{c_k})/T_b, S(\bar{h}_k)/T_b), \quad (3)$$

where  $T_b$  is the temperature parameter,  $S$  is the function to measure SND (detailed in App. B.2),  $h_{c_k} = g_S \circ f_{c_k}$ , and  $\bar{h}_k = g_S \circ \bar{f}_k$ . The initial model for client  $k$  is therefore computed as  $f_k^{\text{init}} = \beta_{k,0} f_{c_k} + \beta_{k,1} \sum_c \alpha_{k,c} \tilde{f}_c$ . Expanding this equation with respect to the cluster model  $f_c$  yields the final cluster weights  $\mathbf{v}_k$ , shown in App. B.3.

On the server side, the global coefficients  $A_{c \rightarrow c'}$  and  $B_{c,i}$  can be estimated by averaging the local weights  $\alpha_{k,c'}$  and  $\beta_{k,i}$  from the previous round within each cluster  $c$  because each cluster comprises clients from similar domains. This ensures the cluster weights in  $f_k^{\text{init}}$  reflect both local preferences and global trends, aligning individual client needs.

Further discussion on privacy concerns and additional costs in distributing all soft cluster models to each client is found in App. D. It also provides a proposal for reducing communication, computation, and storage costs of our approach by shifting client-side computations to the server.

## 4.3. Local Adaptation with Pseudo-Labeling

To adapt the initial model  $f_k^{\text{init}}$  constructed in the previous section to each client  $k$  from only unlabeled data, our local adaptation module adopts the pseudo-labeling method for its simplicity and effectiveness (Fig. 3), while any existing SFDA method could be applicable for local adaptation. This choice is particularly pertinent given the limited computational resources and memory capacities of individual clients, compared with other SFDA methods

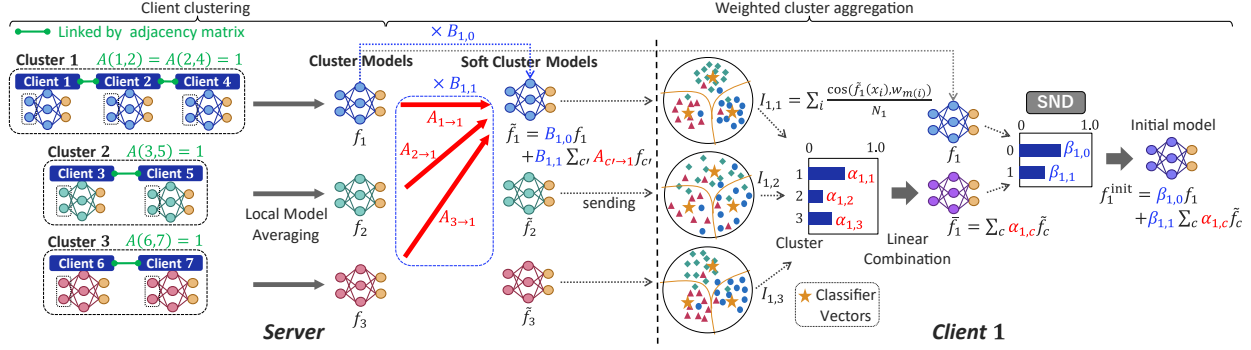


Figure 2. **Weighted cluster aggregation (WCA)**. Our WCA consists of server-side (left) and client-side (right) operations. The client-side operations are represented by client 1 in cluster 1. Firstly, the server forms cluster models by averaging the local models within each cluster and develops soft cluster models by integrating models from other clusters. Next, the client merges these soft cluster models utilizing cluster weights determined by how each feature extractor aligns each sample with the respective classifier vector. Additionally, the client combines the merged model with the corresponding cluster model using Soft Neighborhood Density (SND) weights [41].

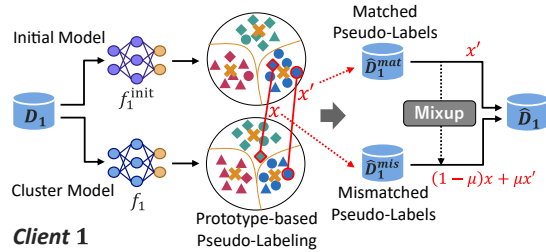


Figure 3. **Pseudo-labeling strategy**. The operation is signified by client 1 in cluster 1. Our strategy involves two phases: prototype-based pseudo-labeling with initial and cluster models, and mixup to generate samples with more reliable pseudo-labels. The first phase selects pseudo-labels closer to prototypes, while the second mixes samples with mismatched and matched pseudo-labels.

[39, 51, 56, 57, 62, 63]. We further alter an existing feature prototype-based pseudo-labeling strategy [6, 29, 40, 54] to lessen overfitting to sparse client data and forgetting global knowledge, and acquire more reliable pseudo-labels. Overfitting is mitigated by using round-fixed pseudo-labels while reliable pseudo-labels are obtained by using both initial and cluster models for pseudo-labeling, followed by mixup.

**Pseudo-labeling strategy.** First, we describe the existing prototype-based pseudo-labeling via a model  $h = g \circ f$ . The procedure involves calculating class-wise feature prototypes  $p_m$  for each class  $m$  and subsequently determining pseudo-labels  $\hat{y}$  for data points  $x$  using the cosine similarity:

$$p_m = \frac{\sum_{x \in D_k} (h(x))_m f(x)}{\sum_{x \in D_k} (h(x))_m}, \quad \hat{y} = \arg \max_m d_m, \quad (4)$$

where  $d_m := \cos(f(x), p_m)$  and  $\cos(\cdot, \cdot)$  represents the cosine similarity. Clients then recompute the prototypes using these one-hot pseudo-labels instead of  $h(x)$  in Eq. (4) and

obtains the final pseudo-labels using these prototypes.

In our setting, each client obtains  $f_k^{\text{init}}$  and  $f_{c_k}$  in each round, allowing every data point  $x$  to have two pseudo-labels generated by these models. We select the pseudo-label with the larger similarity  $d_{\hat{y}} = \cos(f(x), p_{\hat{y}})$ . However, as this similarity is calculated on the feature space constructed by each feature extractor, a similarity normalized by the mean similarity between prototypes  $\sum_{m \neq m'} \cos(p_m, p_{m'}) / (M(M-1))$  is used to select the correct pseudo-label. This results in a pseudo-labelled dataset  $\hat{D}_k = \{x_i, \hat{y}_i\}_{i=1}^{N_k}$  for each client. Contrary to common practices in SFDA, where pseudo-labels are updated in each epoch, our method keep these pseudo-labels constant throughout each round, ensuring the preservation of the global knowledge in  $f_k^{\text{init}}$  and  $f_{c_k}$ .

If  $f_k^{\text{init}}$  and  $f_{c_k}$  produce identical pseudo-labels, these are deemed more reliable; distinct pseudo-labels indicate reduced reliability. Therefore, we divide  $\hat{D}_k$  into  $\hat{D}_k^{\text{mat}}$  and  $\hat{D}_k^{\text{mis}}$ , having matched and mismatched pseudo-labels between  $f_k^{\text{init}}$  and  $f_{c_k}$ , respectively. For each instance  $(x, \hat{y}) \in \hat{D}_k^{\text{mis}}$ , we mix  $x$  with the randomly sampled data point  $(x', \hat{y}) \in \hat{D}_k^{\text{mat}}$ , resulting in  $x_{\text{mix}} = (1 - \mu)x + \mu x'$  with the more reliable pseudo-label  $\hat{y}$ . Here,  $\mu$  is a balancing parameter fixed to 0.55 in our experiments. We thus generate a pseudo-labeled dataset  $\hat{D}_k = \hat{D}_k^{\text{mat}} \cup \hat{D}_k^{\text{mix}}$ , where  $\hat{D}_k^{\text{mix}}$  is a dataset derived from the mixup operation.

**Loss function.** We adopt the same loss function as SHOT [29] because of its simplicity. The total loss function  $\mathcal{L}_k = \mathcal{L}_{\text{IM}} + \lambda \mathcal{L}_{\text{CE}}$  for client  $k$  with a balancing parameter  $\lambda$  comprises the cross-entropy loss for the pseudo-labeled dataset  $\hat{D}_k$ :  $\mathcal{L}_{\text{CE}}$  and an Information Maximization (IM) loss [13, 29] for the unlabeled dataset  $D_k$ :  $\mathcal{L}_{\text{IM}}$  (detailed in App. B.4). The IM loss promotes confident model outputs and discourages trivial score distributions.

Table 2. **Results (Digit-Five, PACS, and Office-Home)**. Each row represents the selected source domain denoted by the initial two letters for each dataset, and the numbers are the mean values  $\pm$  standard deviations of averaged accuracy (%) across all clients and all target domains. These figures are computed from 15 trials. The best numbers in each row are denoted in bold and the second best are underlined.

Datasets	Methods						
	Source Only	Local	FedAvg	LADD	FedPCL+PL	FedWCA (ours)	
Digit-Five	MN	41.81 $\pm$ 0.46	50.59 $\pm$ 1.25	<u>55.88 <math>\pm</math> 2.59</u>	49.00 $\pm$ 0.82	47.84 $\pm$ 0.79	<b>72.74 <math>\pm</math> 3.57</b>
	SV	57.40 $\pm$ 1.39	83.00 $\pm$ 0.99	<u>83.76 <math>\pm</math> 1.97</u>	78.40 $\pm$ 0.64	77.73 $\pm$ 0.56	<b>90.13 <math>\pm</math> 2.08</b>
	MN-M	59.87 $\pm$ 1.71	69.44 $\pm$ 0.97	<u>80.14 <math>\pm</math> 0.46</u>	72.93 $\pm$ 1.25	70.06 $\pm$ 1.94	<b>82.93 <math>\pm</math> 0.79</b>
	US	46.93 $\pm$ 0.54	56.76 $\pm$ 1.13	53.23 $\pm$ 0.73	<u>56.93 <math>\pm</math> 1.59</u>	56.41 $\pm$ 0.62	<b>58.56 <math>\pm</math> 4.26</b>
	SY	68.69 $\pm$ 0.27	81.48 $\pm$ 0.32	80.81 $\pm$ 0.98	<u>83.79 <math>\pm</math> 0.66</u>	81.51 $\pm$ 0.33	<b>84.94 <math>\pm</math> 1.15</b>
	<b>Avg.</b>	54.94 $\pm$ 9.92	68.25 $\pm$ 13.07	<u>70.76 <math>\pm</math> 13.49</u>	67.67 $\pm$ 13.25	66.71 $\pm$ 12.88	<b>77.86 <math>\pm</math> 11.56</b>
PACS	Ar	65.87 $\pm$ 1.19	73.01 $\pm$ 1.40	<u>76.57 <math>\pm</math> 1.38</u>	74.20 $\pm$ 1.48	73.83 $\pm$ 2.84	<b>80.63 <math>\pm</math> 2.44</b>
	Ca	68.66 $\pm$ 1.02	77.27 $\pm$ 1.23	76.15 $\pm$ 0.72	77.37 $\pm$ 1.12	<u>79.18 <math>\pm</math> 1.51</u>	<b>83.18 <math>\pm</math> 0.86</b>
	Ph	54.20 $\pm$ 1.77	63.85 $\pm$ 1.38	60.94 $\pm$ 0.96	64.77 $\pm$ 1.80	<u>65.27 <math>\pm</math> 1.07</u>	<b>65.50 <math>\pm</math> 1.24</b>
	Sk	50.07 $\pm$ 3.21	75.56 $\pm$ 6.05	<u>80.42 <math>\pm</math> 6.63</u>	61.82 $\pm$ 4.88	59.41 $\pm$ 6.22	<b>84.22 <math>\pm</math> 6.60</b>
	<b>Avg.</b>	59.70 $\pm$ 8.18	72.42 $\pm$ 6.10	<u>73.52 <math>\pm</math> 8.23</u>	69.54 $\pm$ 7.02	69.42 $\pm$ 8.42	<b>78.38 <math>\pm</math> 8.38</b>
Office-Home	Ar	54.03 $\pm$ 0.30	62.50 $\pm$ 0.78	<u>63.81 <math>\pm</math> 0.85</u>	57.33 $\pm$ 0.72	58.20 $\pm$ 0.73	<b>66.06 <math>\pm</math> 0.50</b>
	Cl	58.36 $\pm$ 1.29	64.57 $\pm$ 0.70	<b>68.73 <math>\pm</math> 0.75</b>	61.72 $\pm$ 0.61	62.39 $\pm$ 0.98	<u>68.32 <math>\pm</math> 0.42</u>
	Pr	55.36 $\pm$ 0.72	59.17 $\pm$ 0.40	60.15 $\pm$ 0.48	57.22 $\pm$ 0.76	58.85 $\pm$ 0.46	<b>61.46 <math>\pm</math> 0.88</b>
	Re	62.28 $\pm$ 0.35	65.38 $\pm$ 1.15	<u>65.93 <math>\pm</math> 0.34</u>	64.23 $\pm$ 0.44	65.75 $\pm$ 0.80	<b>68.06 <math>\pm</math> 0.46</b>
<b>Avg.</b>	57.51 $\pm$ 3.33	62.90 $\pm$ 2.54	<u>64.65 <math>\pm</math> 3.22</u>	60.25 $\pm$ 2.96	61.30 $\pm$ 3.14	<b>65.97 <math>\pm</math> 2.83</b>	

## 5. Experiment

**Datasets.** Our method was evaluated on three classification datasets, each with diverse domains. The **Digit-Five** dataset encompasses 5 benchmarks for handwritten-digit classification across 10 categories. It includes MNIST [23], MNIST-M [11], SVHN [36], SYNTH [11], and USPS. The **PACS** [25] dataset consists of 4 domains with 7 categories each, including Art Painting, Cartoon, Photo, and Sketch. The **Office-Home** [53] dataset is designed for object recognition, comprising 4 domains with 65 categories each and includes Art, Clipart, Product, and Real-World. For each dataset, one domain was selected as the source, and the remaining domains served as target domains. Data samples from each target domain were distributed to 8 clients for Digit-Five and 3 clients for PACS and Office-Home per domain, all in an i.i.d. manner. This resulted in a total of 32 clients for Digit-Five and 9 for PACS and Office-Home.

**Models.** Our model architecture reflects common structures used in SFDA research like SHOT [29], incorporating a feature extractor, a dimension-reducing bottleneck layer, and a classifier. The feature extractor differs across datasets. Following prior domain adaptation and FL studies, we use a randomly initialized LeNet [23] with two convolution layers for Digit-Five. For PACS and Office-Home, we use ResNet-18 and ResNet-50 [15] respectively, both pre-trained on ImageNet. The classifier is the last linear layer and the bottleneck layer consists of a fully connected layer equipped with batch normalization [18] and dropout [46] for all datasets.

**Baselines.** For comparative analysis, we benchmark our method **FedWCA** against several baseline models. **Source Only** uses only pre-trained source models. In terms of

SFDA approaches, **Local** and **FedAvg** apply SFDA methods to client-side learning. While ‘Local’ represents a non-federated approach, ‘FedAvg’ incorporates model aggregation by FedAvg [34] at the server. As a foundational SFDA method, SHOT [29] is used in these approaches, with round-fixed pseudo-labels for FedAvg, akin to our method. We also assess against **LADD** [45], an existing FFREEDA method. To examine if existing FL approaches for non-i.i.d. data are applicable to FFREEDA by adding a pseudo-labeling step, we compare our method with **FedPCL+PL**. This method integrates FedPCL [50] with prototype-based pseudo-labeling. FedPCL aligns with our setting as it adapts pre-trained models to clients. See App. E.3 for comparisons with other FL methods extended to FFREEDA.

**Evaluation.** We trained five source models using standard cross-entropy loss without incorporating any target data information to ensure a fair comparison. We then conducted three trials for each method per source model, totaling fifteen trials. The evaluation metric is the average accuracy across all clients, assessed by the final personalized models using client-specific test data. The results provide the mean values of this average accuracy over all trials.

See App. C for more detailed experimental settings.

### 5.1. Comparison with Baselines

The comparison across the three datasets (Tab. 2) reveals varied performances. The ‘Source Only’ approach demonstrates limited effectiveness, while the ‘Local’ approach, applying SFDA to local learning, significantly enhances performance. ‘FedAvg’ further improves upon ‘Local’, showcasing the benefits of FL. However, ‘FedAvg’ ex-

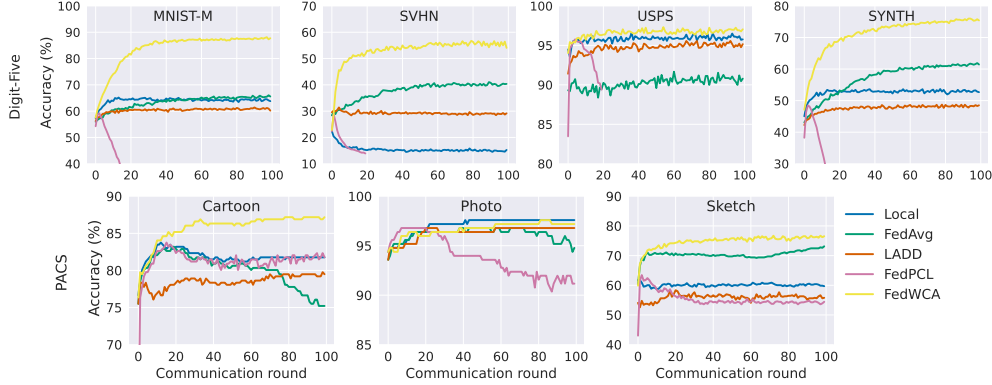


Figure 4. **Comparison of average accuracy over clients within each target domain.** The source domain is MNIST for Digit-Five (top) and Art Painting for PACS (bottom). The target domains are the others. FedWCA considerably enhances the accuracy of clients in specific target domains such as SVHN for Digit-Five and Sketch for PACS, while maintaining the performance of other domains.

periences performance drops in some cases due to domain shift, such as a decrease from 56.76% to 53.23% on USPS for Digit-Five. ‘LADD’ and ‘FedPCL+PL’ mitigate domain shift issues somewhat and improve accuracy in some source domains, but are inferior to ‘FedAvg’ in many domains. This is because LADD’s primary focus is not classification tasks and therefore it uses a simple pseudo-labelling strategy, and furthermore it does not fully utilize knowledge from all domains. On the other hand, FedPCL originally employs true labels at every step and substituting them all with pseudo-labels could compromise performance.

Our FedWCA outperforms the best compared methods by notable margins on almost all source domains (7.10% for Digit-Five, 4.86% for PACS, 1.32% for Office-Home on average). The results also confirm that FedWCA statistically outperforms the others in a majority of source domains. The improvement on Office-Home is more modest, likely due to the smaller domain gap in this dataset, which can be judged from the fact that the improvement for ‘Source Only’ by ‘Local’ is considerably smaller (5.39%) than for the other datasets (13.31% for Digit-Five and 12.72% for PACS).

In Fig. 4, we analyze the changes in average accuracy of clients in each target domain, using specific source domains, such as MNIST for Digit-Five (top) and Art Painting for PACS (bottom). Unlike ‘FedAvg’, FedWCA significantly improves accuracy in almost all target domains without compromising accuracy in specific target domains such as USPS for Digit-Five and Photo for PACS.

## 5.2. Ablation Studies

### 5.2.1 Cluster Weight Calculation

To validate the effectiveness of our cluster weight calculation method, we conducted an ablation study comparing our method with simpler cluster weight computation methods including: (1) **One-hot**: Assigns a weight of 1 to each

Table 3. **Ablation on cluster weights calculation.** The FedWCA’s cluster weights calculation is replaced by various methods such as one-hot weights (One-hot), equal weights (Equal), a hybrid of them (One-Equal), an adaptive version of One-Equal (One-Equal-A), and a local calculation-only FedWCA variant described in Sec. 4.2 (FedWCA-L). The original FedWCA surpasses all other methods across all PACS source domains.

Methods	PACS				
	Ar	Ca	Ph	Sk	Avg.
One-hot	73.22	77.34	64.45	80.40	73.85
Equal	76.91	76.03	61.16	79.55	73.41
One-Equal	77.89	79.69	64.76	83.13	76.37
One-Equal-A	77.80	79.17	63.64	83.29	75.98
FedWCA-L	77.12	81.94	62.25	82.31	75.91
FedWCA	<b>80.63</b>	<b>83.18</b>	<b>65.50</b>	<b>84.22</b>	<b>78.38</b>

client’s own cluster and 0 to others. (2) **Equal**: Allocates an equal weight to all clusters, resembling FedAvg. (3) **One-Equal**: A hybrid approach that assigns a weight of  $p$  (set to 0.8) to each client’s cluster and distributes the remaining weight equally among other clusters. (4) **One-Equal-A**: An adaptive method that combines models created by One-hot and Equal using SND, as in FedWCA. (5) **FedWCA-L**: A variant of FedWCA wherein cluster weights are computed solely locally, described in Sec. 4.2.

The performance comparison of these methods on PACS is shown in Tab. 3. The hybrid approaches, One-Equal and its adaptive variant, and FedWCA-L, increase accuracy compared to simpler methods like One-hot and Equal. FedWCA, however, outperforms them across all source domains. The superior performance illustrates the benefits of our method that integrates global and local insights in cluster weight calculation. See App. E.1 for examples of specific cluster weights in our methods.

Table 4. **Ablation on layers used for clustering.** Art Painting of PACS is used as the source domain, and ResNet-18 is employed. Clients  $C, C', C''$  belong to Cartoon,  $P, P', P''$  to Photo, and  $S, S', S''$  to Sketch. The table below reports the cluster IDs (denoted by  $\circ, \triangle$ , and  $\square$ ) assigned to each client. Ideally, each client group ( $C, C', C''$ ), ( $P, P', P''$ ), and ( $S, S', S''$ ) should be grouped together, as is achieved by the first and second layers.

Layer	Clients								
	$C$	$C'$	$C''$	$P$	$P'$	$P''$	$S$	$S'$	$S''$
1st	$\circ$	$\circ$	$\circ$	$\triangle$	$\triangle$	$\triangle$	$\square$	$\square$	$\square$
2nd	$\circ$	$\circ$	$\circ$	$\triangle$	$\triangle$	$\triangle$	$\square$	$\square$	$\square$
3rd	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$
4th	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\triangle$	$\triangle$	$\triangle$
17th	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$
All	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$	$\circ$

### 5.2.2 Client Clustering Method

Tab. 4 evaluates the clustering performance using different layers’ parameters in ResNet-18, showing that using only the first or second layer is the best practice (see App. E.2 for full results). The results indicate that clustering performance diminishes as a deeper layer rather than the first layer (*e.g.*, 17th) is used. Using all layers leads to a single cluster (indexed by  $\circ$  in Tab. 4) due to the curse of dimensionality.

Despite successful clustering in PACS and Office-Home, some errors were observed in Digit-Five (see App. E.2). Nonetheless, our method improves performance even with incorrect cluster assignments by effectively combining cluster models with individualized weights (App. E.2).

### 5.2.3 Pseudo-labeling Strategy

We examine the contribution of each key component of our pseudo-labeling strategy. The pseudo-labeling strategy of FedWCA consists of four key components (see Sec. 4.3): prototype-based pseudo-labeling (PP), fixed pseudo-labels (FP) in each round, two-models-based pseudo-labeling (TM), which utilizes both the initial and original cluster models, and mixup (MU) between the samples with the matched and mismatched pseudo-labels.

To assess the effectiveness of each component, we incrementally activated them in FedWCA and observed their impact on performance in the PACS dataset. The results, presented in Tab. 5, reveal that the full integration of all components leads to the highest performance across all source domains. This finding underscores the significance of each element, with every component contributing meaningfully to the overall efficacy of the method.

The above results demonstrate that PP and FP independently enhance accuracy and may also be incorporated into LADD. Similarly, the IM loss adopted by our method could contribute to LADD’s expansion, tailoring it more towards the classification task, akin to the local adaptation module

Table 5. **Ablation on pseudo-labeling strategy.** Key components of FedWCA are **PP**: prototype-based pseudo-labeling, **FP**: fixed pseudo-labels in each round, **TM**: two-models-based pseudo-labeling using initial and cluster models, and **MU**: mixup. All components prove effective in most PACS source domains.

PP	FP	TM	MU	PACS				
				Ar	Ca	Ph	Sk	Avg.
	$\checkmark$			72.98	76.44	61.78	68.73	70.29
$\checkmark$				77.63	75.66	65.48	81.31	74.39
$\checkmark$	$\checkmark$			78.37	81.54	65.23	83.56	77.17
$\checkmark$	$\checkmark$	$\checkmark$		78.82	81.72	<b>66.27</b>	83.92	77.68
$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	<b>80.63</b>	<b>83.18</b>	65.50	<b>84.22</b>	<b>78.38</b>

Table 6. **Comparison with extended LADD.** The prior FFREEDA method LADD is extended by incorporating **PP**: prototype-based pseudo-labeling, **FP**: fixed pseudo-labels in each round, and **IM**: IM loss. Nevertheless, FedWCA outperforms it.

Datasets	Methods	LADD+PP+FP+IM		FedWCA
Digit-Five	MN	60.45 $\pm$ 2.14		<b>72.74 <math>\pm</math> 3.57</b>
	SV	90.06 $\pm$ 2.41		<b>90.13 <math>\pm</math> 2.08</b>
	MN-M	72.11 $\pm$ 3.06		<b>82.93 <math>\pm</math> 0.79</b>
	US	56.96 $\pm$ 1.67		<b>58.56 <math>\pm</math> 4.26</b>
	SY	<b>87.12 <math>\pm</math> 0.65</b>		84.94 $\pm$ 1.15
	Avg.	73.34 $\pm$ 13.71		<b>77.86 <math>\pm</math> 11.56</b>
PACS	Ar	73.19 $\pm$ 1.73		<b>80.63 <math>\pm</math> 2.44</b>
	Ca	75.98 $\pm$ 2.49		<b>83.18 <math>\pm</math> 0.86</b>
	Ph	65.15 $\pm$ 1.27		<b>65.50 <math>\pm</math> 1.24</b>
	Sk	79.24 $\pm$ 6.76		<b>84.22 <math>\pm</math> 6.60</b>
	Avg.	73.39 $\pm$ 6.41		<b>78.38 <math>\pm</math> 8.38</b>
Office-Home	Ar	65.04 $\pm$ 0.82		<b>66.06 <math>\pm</math> 0.50</b>
	Cl	67.88 $\pm$ 0.67		<b>68.32 <math>\pm</math> 0.42</b>
	Pr	61.06 $\pm$ 0.53		<b>61.46 <math>\pm</math> 0.88</b>
	Re	67.58 $\pm$ 0.80		<b>68.06 <math>\pm</math> 0.46</b>
Avg.	65.39 $\pm$ 3.14		<b>65.97 <math>\pm</math> 2.83</b>	

with ‘FedAvg’ in our experiments. Tab. 6 compares the performance of the extended LADD with our FedWCA. A comparison between Tab. 2 and Tab. 6 reveals a significant performance boost in the extended LADD over the original. However, FedWCA outperforms this improved LADD, further supporting our method’s efficiency.

## 6. Conclusion

This paper introduces FedWCA, a novel method for FFREEDA classification, addressing two problems: lack of labeled data post-pre-training and domain shift among clients. Experimental results show FedWCA’s superior performance over several baselines, validating its effectiveness in FFREEDA. We discuss the limitations in App. D.

## Acknowledgments

This R&D includes the results of ”Research and development of optimized AI technology by secure data coordination (JPMI00316)” by the Ministry of Internal Affairs and Communications (MIC), Japan.



## References

- [1] Christopher Briggs, Zhong Fan, and Peter Andras. Federated learning with hierarchical clustering of local updates to improve training on non-iid data. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9, 2020. [2](#)
- [2] Debora Caldarola, Massimiliano Mancini, Fabio Galasso, Marco Ciccone, Emanuele Rodolà, and Barbara Caputo. Cluster-driven graph federated learning over multiple domains. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2743–2752, 2021. [2](#)
- [3] Jichan Chung, Kangwook Lee, and Kannan Ramchandran. Federated unsupervised clustering with generative models. In *AAAI 2022 International Workshop on Trustable, Verifiable and Auditable Federated Learning*, 2022. [3](#)
- [4] Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting shared representations for personalized federated learning. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 2089–2099. PMLR, 18–24 Jul 2021. [2, 3](#)
- [5] Enmao Diao, Jie Ding, and Vahid Tarokh. Semifl: Semi-supervised federated learning for unlabeled clients with alternate training. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 17871–17884. Curran Associates, Inc., 2022. [2, 3](#)
- [6] Ning Ding, Yixing Xu, Yehui Tang, Chao Xu, Yunhe Wang, and Dacheng Tao. Source-free domain adaptation via distribution estimation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7202–7212, 2022. [2, 3, 5](#)
- [7] Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar. Personalized federated learning with theoretical guarantees: a model-agnostic meta-learning approach. In *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS’20*, Red Hook, NY, USA, 2020. Curran Associates Inc. [2, 3](#)
- [8] Chenyou Fan, Junjie Hu, and Jianwei Huang. Private semi-supervised federated learning. In Lud De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 2009–2015. International Joint Conferences on Artificial Intelligence Organization, 7 2022. Main Track. [3](#)
- [9] Haozhe Feng, Zhaoyang You, Minghao Chen, Tianye Zhang, Minfeng Zhu, Fei Wu, Chao Wu, and Wei Chen. Kd3a: Unsupervised multi-source decentralized domain adaptation via knowledge distillation. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 3274–3283. PMLR, 18–24 Jul 2021. [2, 3](#)
- [10] Shaoduo Gan, Akhil Mathur, Anton Isopoussu, Fahim Kawsar, Nadia Berthouze, and Nicholas D. Lane. Fruda: Framework for distributed adversarial domain adaptation. *IEEE Transactions on Parallel and Distributed Systems*, 33(11):3153–3164, 2022. [2, 3](#)
- [11] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. *Domain-Adversarial Training of Neural Networks*, pages 189–209. Springer International Publishing, Cham, 2017. [6](#)
- [12] Avishek Ghosh, Jichan Chung, Dong Yin, and Kannan Ramchandran. An efficient framework for clustered federated learning. In *Proceedings of the 34th International Conference on Neural Information Processing Systems, NIPS’20*, Red Hook, NY, USA, 2020. Curran Associates Inc. [2, 17](#)
- [13] Ryan Gomes, Andreas Krause, and Pietro Perona. Discriminative clustering by regularized information maximization. In *Proceedings of the 23rd International Conference on Neural Information Processing Systems - Volume 1, NIPS’10*, page 775–783, Red Hook, NY, USA, 2010. Curran Associates Inc. [5, 12](#)
- [14] Andrey Gromov, Kushal Tirumala, Hassan Shapourian, Paolo Glorioso, and Daniel A. Roberts. The unreasonable ineffectiveness of the deeper layers, 2024. [3](#)
- [15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. [6](#)
- [16] Junyuan Hong, Zhuangdi Zhu, Shuyang Yu, Zhangyang Wang, Hiroko H. Dodge, and Jiayu Zhou. Federated adversarial debiasing for fair and transferable representations. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, KDD ’21*, page 617–627, New York, NY, USA, 2021. Association for Computing Machinery. [2, 3](#)
- [17] Yutao Huang, Lingyang Chu, Zirui Zhou, Lanjun Wang, Jiangchuan Liu, Jian Pei, and Yong Zhang. Personalized cross-silo federated learning on non-iid data. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(9):7865–7873, May 2021. [2, 3, 17](#)
- [18] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France, 07–09 Jul 2015. PMLR. [6](#)
- [19] Wonyong Jeong, Jaehong Yoon, Eunho Yang, and Sung Ju Hwang. Federated semi-supervised learning with inter-client consistency & disjoint learning. In *International Conference on Learning Representations*, 2021. [2, 3](#)
- [20] Hua Kang, Zhiyang Li, and Qian Zhang. Communicational and computational efficient federated domain adaptation. *IEEE Transactions on Parallel and Distributed Systems*, 33(12):3678–3689, 2022. [2, 3](#)
- [21] Hansol Kim, Youngjun Kwak, Minyoung Jung, Jinho Shin, Youngsung Kim, and Changick Kim. Protofl: Unsupervised federated learning via prototypical distillation. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6447–6456, 2023. [3](#)

- [22] Kavya Koppurapu and Eric Lin. Fedfmc: Sequential efficient federated learning on non-iid data, 2020. [2](#)
- [23] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998. [6](#)
- [24] Jonghyun Lee, Dahun Jung, Junho Yim, and Sungroh Yoon. Confidence score for source-free unsupervised domain adaptation. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 12365–12377. PMLR, 17–23 Jul 2022. [2](#)
- [25] Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M. Hospedales. Deeper, broader and artier domain generalization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5543–5551, 2017. [6](#)
- [26] Ming Li, Qingli Li, and Yan Wang. Class balanced adaptive pseudo labeling for federated semi-supervised learning. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 16292–16301, 2023. [2, 3](#)
- [27] Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. In I. Dhillon, D. Papailiopoulos, and V. Sze, editors, *Proceedings of Machine Learning and Systems*, volume 2, pages 429–450, 2020. [2, 17](#)
- [28] Zexi Li, Jiaxun Lu, Shuang Luo, Didi Zhu, Yunfeng Shao, Yinchuan Li, Zhimeng Zhang, Yongheng Wang, and Chao Wu. Towards effective clustered federated learning: A peer-to-peer framework with adaptive neighbor matching. *IEEE Transactions on Big Data*, pages 1–16, 2022. [2](#)
- [29] Jian Liang, Dapeng Hu, and Jiashi Feng. Do we really need to access the source data? Source hypothesis transfer for unsupervised domain adaptation. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 6028–6039. PMLR, 13–18 Jul 2020. [2, 3, 5, 6, 12, 13, 14](#)
- [30] Xiaoxiao Liang, Yiqun Lin, Huazhu Fu, Lei Zhu, and Xiaomeng Li. Rscfed: Random sampling consensus federated semi-supervised learning. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10144–10153, 2022. [3](#)
- [31] Mattia Litrico, Alessio Del Bue, and Pietro Morerio. Guiding pseudo-labels with uncertainty estimation for source-free unsupervised domain adaptation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7640–7650, 2023. [2](#)
- [32] Ekdeep Lubana, Chi Ian Tang, Fahim Kawsar, Robert Dick, and Akhil Mathur. Orchestra: Unsupervised federated learning via globally consistent clustering. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 14461–14484. PMLR, 17–23 Jul 2022. [3](#)
- [33] Yishay Mansour, Mehryar Mohri, Jae Ro, and Ananda Theertha Suresh. Three approaches for personalization with applications to federated learning. *arXiv preprint arXiv:2002.10619*, 2020. [2, 3](#)
- [34] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Aarti Singh and Jerry Zhu, editors, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, volume 54 of *Proceedings of Machine Learning Research*, pages 1273–1282. PMLR, 20–22 Apr 2017. [6](#)
- [35] Junki Mori, Tomoyuki Yoshiyama, Ryo Furukawa, and Isamu Teranishi. Personalized federated learning with multi-branch architecture. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2023. [2, 3](#)
- [36] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bisaccho, Bo Wu, and Andrew Y. Ng. Reading digits in natural images with unsupervised feature learning. In *NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011*, 2011. [6](#)
- [37] Chamath Palihawadana, Nirmalie Wiratunga, Anjana Wijekoon, and Harsha Kalutarage. FedSim: Similarity guided model aggregation for federated learning. *Neurocomput.*, 483(C):432–445, apr 2022. [2](#)
- [38] Xingchao Peng, Zijun Huang, Yizhe Zhu, and Kate Saenko. Federated adversarial domain adaptation. In *International Conference on Learning Representations*, 2020. [2, 3](#)
- [39] Zhen Qiu, Yifan Zhang, Hongbin Lin, Shuaicheng Niu, Yanxia Liu, Qing Du, and Mingkui Tan. Source-free domain adaptation via avatar prototype generation and adaptation. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 2921–2927. International Joint Conferences on Artificial Intelligence Organization, 8 2021. Main Track. [5](#)
- [40] Sanqing Qu, Guang Chen, Jing Zhang, Zhijun Li, Wei He, and Dacheng Tao. BMD: A general class-balanced multicentric dynamic prototype strategy for source-free domain adaptation. In Shai Avidan, Gabriel Brostow, Moustapha Cissé, Giovanni Maria Farinella, and Tal Hassner, editors, *Computer Vision – ECCV 2022*, pages 165–182, Cham, 2022. Springer Nature Switzerland. [2, 3, 5](#)
- [41] Kuniaki Saito, Donghyun Kim, Piotr Teterwak, Stan Sclaroff, Trevor Darrell, and Kate Saenko. Tune it the right way: Unsupervised validation of domain adaptation via soft neighborhood density. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9164–9173, 2021. [4, 5, 12, 13](#)
- [42] Saquib Sarfraz, Vivek Sharma, and Rainer Stiefelhagen. Efficient parameter-free clustering using first neighbor relations. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8926–8935, 2019. [3](#)
- [43] Felix Sattler, Klaus-Robert Müller, and Wojciech Samek. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. *IEEE Transactions on Neural Networks and Learning Systems*, 32(8):3710–3722, 2021. [2](#)
- [44] Mykola Servetnyk, Carrson C. Fung, and Zhu Han. Unsupervised federated learning for unbalanced data. In *GLOBE-*

- COM 2020 - 2020 IEEE Global Communications Conference, pages 1–6, 2020. 3
- [45] Donald Shenaj, Eros Fani, Marco Toldo, Debora Caldarola, Antonio Tavera, Umberto Michieli, Marco Ciccone, Pietro Zanuttigh, and Barbara Caputo. Learning across domains and devices: Style-driven source-free domain adaptation in clustered federated learning. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 444–454, 2023. 1, 2, 3, 6
- [46] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958, 2014. 6
- [47] Benyuan Sun, Hongxing Huo, YI YANG, and Bo Bai. Partialfed: Cross-domain personalized federated learning via partial initialization. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 23309–23320. Curran Associates, Inc., 2021. 2, 3
- [48] Yuwei Sun, Ng Chong, and Hideya Ochiai. Feature distribution matching for federated domain generalization. In Emteyaz Khan and Mehmet Gonen, editors, *Proceedings of The 14th Asian Conference on Machine Learning*, volume 189 of *Proceedings of Machine Learning Research*, pages 942–957. PMLR, 12–14 Dec 2023. 2, 3
- [49] Alysa Ziyang Tan, Han Yu, Lizhen Cui, and Qiang Yang. Towards personalized federated learning. *IEEE Transactions on Neural Networks and Learning Systems*, 34(12):9587–9603, 2023. 2
- [50] Yue Tan, Guodong Long, Jie Ma, LU LIU, Tianyi Zhou, and Jing Jiang. Federated learning from pre-trained models: A contrastive learning approach. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 19332–19344. Curran Associates, Inc., 2022. 3, 6
- [51] Jiayi Tian, Jing Zhang, Wen Li, and Dong Xu. Vdm-da: Virtual domain modeling for source data-free domain adaptation. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(6):3749–3760, 2022. 5
- [52] Pu Tian, Weixian Liao, Wei Yu, and Erik Blasch. Wsccl: A weight-similarity-based client clustering approach for non-iid federated learning. *IEEE Internet of Things Journal*, 9(20):20243–20256, 2022. 2
- [53] Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5385–5394, 2017. 6
- [54] Fan Wang, Zhongyi Han, Yongshun Gong, and Yilong Yin. Exploring domain-invariant parameters for source free domain adaptation. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7141–7150, 2022. 2, 5
- [55] Huisi Wu, Baiming Zhang, Cheng Chen, and Jing Qin. Federated semi-supervised medical image segmentation via prototype-based pseudo-labeling and contrastive learning. *IEEE Transactions on Medical Imaging*, 43(2):649–661, 2024. 3
- [56] Shiqi Yang, yaxing wang, Joost van de Weijer, Luis Herranz, and Shangling Jui. Exploiting the intrinsic neighborhood structure for source-free domain adaptation. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 29393–29405. Curran Associates, Inc., 2021. 5
- [57] Shiqi Yang, Yaxing Wang, Kai Wang, SHANGLING JUI, and Joost van de weijer. Attracting and dispersing: A simple approach for source-free domain adaptation. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. 5
- [58] Chun-Han Yao, Boqing Gong, Hang Qi, Yin Cui, Yukun Zhu, and Ming-Hsuan Yang. Federated multi-target domain adaptation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1424–1433, January 2022. 2, 3
- [59] Riccardo Zaccone, Andrea Rizzardi, Debora Caldarola, Marco Ciccone, and Barbara Caputo. Speeding up heterogeneous federated learning with sequentially trained superclients. In *2022 26th International Conference on Pattern Recognition (ICPR)*, pages 3376–3382, 2022. 2
- [60] Fengda Zhang, Kun Kuang, Long Chen, Zhaoyang You, Tao Shen, Jun Xiao, Yin Zhang, Chao Wu, Fei Wu, Yueting Zhuang, et al. Federated unsupervised representation learning. *Frontiers of Information Technology & Electronic Engineering*, 24(8):1181–1193, 2023. 3
- [61] Michael Zhang, Karan Sapra, Sanja Fidler, Serena Yeung, and Jose M. Alvarez. Personalized federated learning with first order model optimization. In *International Conference on Learning Representations*, 2021. 2, 3
- [62] Yixin Zhang, Zilei Wang, and Weinan He. Class relationship embedded learning for source-free unsupervised domain adaptation. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7619–7629, 2023. 5
- [63] Ziyi Zhang, Weikai Chen, Hui Cheng, Zhen Li, Siyuan Li, Liang Lin, and Guanbin Li. Divide and contrast: Source-free domain adaptation via adaptive contrastive learning. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 5137–5149. Curran Associates, Inc., 2022. 5
- [64] Zhengming Zhang, Yaoqing Yang, Zhewei Yao, Yujun Yan, Joseph E. Gonzalez, Kannan Ramchandran, and Michael W. Mahoney. Improving semi-supervised federated learning by reducing the gradient diversity of models. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 1214–1225, 2021. 2, 3