

# FedFSL-CFRD: Personalized Federated Few-Shot Learning with Collaborative Feature Representation Disentanglement

Shanfeng Wang<sup>1,2</sup>, Jianzhao Li<sup>1,3\*</sup>, Zaitian Liu<sup>1,2</sup>, Yourun Zhang<sup>1</sup>, Maoguo Gong<sup>1,4</sup>

<sup>1</sup>Key Laboratory of Collaborative Intelligence Systems, Ministry of Education, Xidian University

<sup>2</sup>School of Cyber Engineering, Xidian University

<sup>3</sup>Guangzhou Institute of Technology, Xidian University

<sup>4</sup>Academy of Artificial Intelligence, College of Mathematics Science, Inner Mongolia Normal University  
sfwang@xidian.edu.cn; lijianzhao@xidian.edu.cn; liuzaitian@stu.xidian.edu.cn; zhangyourun@stu.xidian.edu.cn; gong@ieee.org

## Abstract

Federated few-shot learning (FedFSL) aims to enable the clients to obtain personalized generalization models for unseen categories with only a small number of referenceable samples in the distributed collaborative training paradigm. Most existing FedFSL-related algorithms suffer from domain bias and feature coupling in the presence of data heterogeneity and sample scarcity. In this work, we propose a collaborative feature representation disentanglement (CFRD) scheme for FedFSL to address these issues. After each client receives the global aggregation parameters, the original feature representation is decoupled into global communal features and local personality features with personalized bias representation, to maintain both global consistency and local relevance in the first feature representation disentanglement. On the few-shot metric space about the second feature representation disentanglement, category-independent information is encoded by class-specific and class-irrelevant reconstructions to separate the discriminative features. The proposed scheme collaboratively accomplishes global domain bias feature disentanglement and local category degradation feature disentanglement from client-wise and class-wise. Extensive experiments on three few-shot benchmark datasets conforming to the FedFSL paradigm demonstrate that our proposed method outperforms state-of-the-art approaches in both global generality and local specificity.

## Introduction

Federated learning (FL) is to achieve efficient distributed training among multiple parties or computing nodes while protecting information security, terminal data and personal privacy with model parameters or gradients exchange (Li, He, and Song 2021). Most of the early prevalent FL methods are derived from the FedAvg (McMahan et al. 2017), with the aim of training a global model that performs well on the majority of clients (Mohri, Sivek, and Suresh 2019). However, with the development of federated learning theories and applications, the problems of poor convergence in high data heterogeneity and lack of personalized solutions emerge (Tan et al. 2022a; Huang et al. 2023).

\*Corresponding author.

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

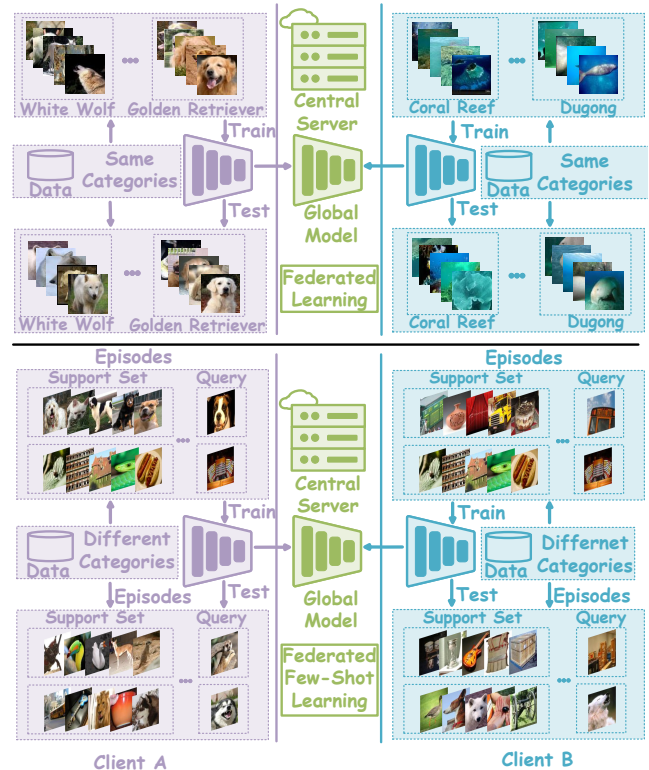


Figure 1: The difference between general federated learning and federated few-shot learning (FedFSL). In the FedFSL, the training and testing categories do not overlap in each client, and the local models are designed to learn from an extremely small number of labeled samples and make effective predictions in novel tasks or domains.

Recently, researchers have delved into more robust heterogeneous FL in terms of global model personalization and training personalized local models (Sun et al. 2020). Specifically, the global model personalization mitigates the client drift problem by reducing statistical heterogeneity between client data, such as data augmentation (Jeong et al. 2018), self-balancing data imbalance (Duan et al. 2020), etc. Alternatively, client-specific personalization improves the adaptive performance to local data by collaboratively learning a

robust global model, such as using a proximal term for local optimization in FedProx (Li et al. 2020), distribution alignment in MOON (Li, He, and Song 2021), sample weight adjustment in FedNova (Wang et al. 2020), meta-learning for fast local adaptation in PerFedAvg (Fallah, Mokhtari, and Ozdaglar 2020), etc. Besides, parameter decoupling for global and local models (Zhang et al. 2021b, 2023b), and knowledge distillation for global to local parameters (Zhu, Hong, and Zhou 2021; Lin et al. 2020) are investigated to obtain personalized model that is more consistent with local data distributions.

The above studies achieve encouraging results for personalized collaborative training, but the generalizability of the local models for novel categories is achieved by task-specific transfer of the global model with strict robustness requirement (Zhao et al. 2022). This is extremely difficult for clients with only a small amount of data, because the tiny percentage of data scale makes it difficult to accurately adapt global parameters in heterogeneity. Consequently, how to enable the local personalization models to achieve fast generalization of unseen categories with only a small number of samples, is the specific motivation for federated few-shot learning (FedFSL) (See Figure 1). Most existing FedFSL-related approaches only focus on personalized federated learning for different few-shot task-specific adaptations, such as facial expression recognition (Shome and Kar 2021), model lightweight (Sun, Yang, and Zhao 2022) and machine fault diagnosis (Zhang et al. 2021a). While in cases of data heterogeneity and sample scarcity, the domain bias and feature coupling have become the difficulties of FedFSL. In fact, this mainly stems from the challenges posed by the complexity of the client and category feature spaces. On the one hand, each client has its own unique local feature space due to the fact that the data is dispersed across different sources in FL. On the other hand, the metric network in FSL is to obtain a more discriminative feature space, such that samples of the same category are more tightly clustered together, while samples of different categories are more widely dispersed.

Given the above insights, we attempt to collaboratively implement feature decoupling at both the client-wise and class-wise, and propose a collaborative feature representation disentanglement (CFRD). Once each client receives the global aggregation parameters, the original feature representation is decoupled into global features and local personalized features with individualized bias representations, to maintain global consistency and local relevance in the first feature representation decoupling. On the few-shot metric spaces regarding the second feature representation decoupling, the class-relevance reconstruction branch is designed on the basis of the few-shot branch to separate discriminative features to realize both class-specific and class-independent reconstruction. Our contributions can be summarized as follows:

- We propose the client-wise feature representation disentanglement to perform global domain bias for FedFS, to achieve a balance between the generality of the global parameter and the personalization of the local parameters, thus alleviating the domain bias problem in FedFSL.

- We propose the class-wise feature representation disentanglement to decouple recognizable features from class-independent information with class-specific and class-independent feature separation, which can effectively solve the category degradation problem in FedFSL scenarios.
- We validate the effectiveness of FedFSL-CFRD on prominent FedFSL benchmarks, showing its superiority and efficiency in feature representation.

## Related Work

### Personalized Federated Learning

Personalized federated learning (PFL) is to improve performance and applicability on non-independently identically-distributed (non-IID) data by personalizing each client with personalized tuning based on the global model or learning personalized models in the collaborative training process, aiming to provide a more accurate model for each client (Wang et al. 2024; Collins et al. 2021). Specifically, methods for local personalized fine-tuning of the global model include PerFedAvg (Fallah, Mokhtari, and Ozdaglar 2020), FedRep (Collins et al. 2021), etc. Approaches for performing targeted training on the global model to obtain additional personalized models include pFedMe (T Dinh, Tran, and Nguyen 2020), Ditto (Li et al. 2021), etc. The work on obtaining local models with personalized aggregation strategies have FedFomo (Zhang et al. 2021b), Partialfed (Sun et al. 2021) FedALA (Zhang et al. 2023b), etc.

### Few-Shot Learning

Few-shot learning (FSL) refers to the exploration of existing knowledge (e.g., pre-trained models or meta-learning methodologies) to enable effective learning and generalization of models to novel classes with a small amount of data (Sung et al. 2018; Sun et al. 2019). The existing related researches are mainly based on meta-learning (Chen et al. 2021; Jamal and Qi 2019), transfer learning (Sun et al. 2019; Yu et al. 2020), and metric learning (Vinyals et al. 2016; Snell, Swersky, and Zemel 2017; Sung et al. 2018) paradigms. There is still relatively little relevant FedFSL research, mainly focusing on the model replacement (Zhao et al. 2022) or the application of PFL to specific FSL task adaptation (Shome and Kar 2021; Sun, Yang, and Zhao 2022; Zhang et al. 2021a).

### Feature Representation Disentanglement

Feature representation disentanglement aims to decompose the representation of complex data into mutually independent or interpretable features. Existing related researches in FL mainly focus on decoupling between different clients, such as feature disentanglement for different modalities (Chen and Zhang 2024) and global model (Yang et al. 2023). In terms of FSL, feature disentanglement to obtain knowledge distillation information (Li et al. 2023), to obtain more discriminative features (Lin et al. 2021; Cheng et al. 2023), and to improve the quality and diversity of generated samples (Dang et al. 2024) have been investigated.

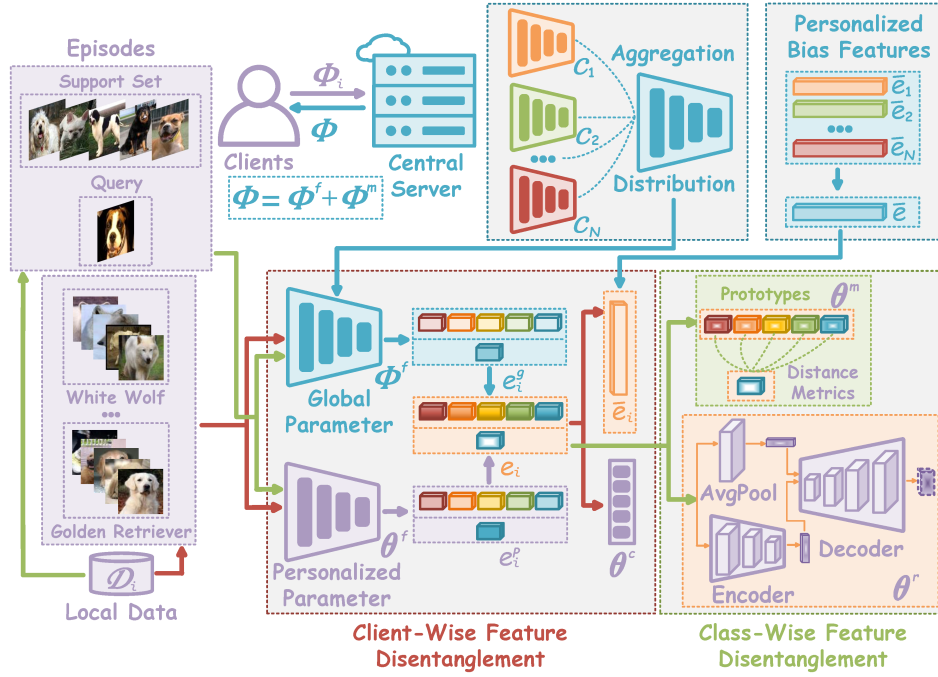


Figure 2: The specific schematic of the FedFSL-CFRD with two components client-wise feature disentanglement and class-wise feature disentanglement. In the FedFSL-CFRD, the discrimination of encoder class-irrelevance is first realized by decoupling the global and local feature representations, as shown in the red stream. Then the category decoupling is realized to further enhance the discrimination of class-relevance, as shown in the green stream.

## Methodology

### FedFSL-CFRD Framework

In FedFSL, it is assumed that a total of  $M$  clients with their private data  $\mathcal{D}_i$  participate in learning collaboratively, where  $\mathcal{D}_i$  for each client  $c_i$  consists of the training set  $\mathcal{D}_i^{base}$  and the test set  $\mathcal{D}_i^{novel}$ . For all participating clients, the categories of the training and test sets do not overlap, *i.e.*,  $(\mathcal{D}_i^{base} \cap \mathcal{D}_i^{novel} = \emptyset)$ . Consistent with the commonly used episodic training (Qiao et al. 2019; Laenen and Bertinetto 2021) in FSL,  $\mathcal{D}_i^{train}$  and  $\mathcal{D}_i^{test}$  are composed of multiple meta-tasks (also known as episodes), where each episode contains a support set  $\mathcal{S}$  and a query set  $\mathcal{Q}$ . In the  $N$ -way  $K$ -shot setting, the support set  $\mathcal{S}$  is composed of  $K$  labeled samples for each of the  $N$  randomly selected categories from the dataset, usually the number of  $K$  is very small, while the query set is composed of a certain number of unlabeled samples drawn from within these  $N$  categories. For each client  $c_i$ , the loss function can be expressed as  $\mathcal{L}_i = \mathcal{L}(\theta_i; \mathcal{D}_i)$ , where  $\theta_i$  represents the local parameters. In FSL,  $\theta_i$  usually contains a feature encoder parameter  $\theta_i^f$  and a metric parameter  $\theta_i^m$ .

In the general FL, all the participants aim to train collaboratively to obtain a global model  $\phi$  with the optimization problem:  $\min_{\phi} \mathbb{E}_{i \in [M]} \{\mathcal{L}_{\mathcal{D}_i}(\phi)\}$ . However, in addition to collaboratively learning for the generalized global model  $\phi^*$ , the optimization objective of FedFSL is also to expect each participating client to obtain the personalized parameter that is more tailored to the lo-

cal data distribution in collaborative learning. Therefore, the optimization problem of FedFSL can be expressed as:  $\min_{\{\theta_1, \dots, \theta_M\}} \mathbb{E}_{i \in [M]} \{\mathcal{L}_{\mathcal{D}_i}(\theta_i)\}$ , which aims to obtain the local personalized parameter:  $\{\theta_1^*, \dots, \theta_M^*\} = \arg \min_{\{\theta_1^*, \dots, \theta_M^*\}} \sum_{i=1}^M \mathcal{L}_i(\theta_i; \mathcal{D}_i)$ .

The specific schematic of the FedFSL-CFRD with two components client-wise feature disentanglement and class-wise feature disentanglement is shown in Figure 2. The overall collaborative learning paradigm of FedFSL-CFRD follows FL, and the local model update follows the episodic training paradigm of FSL. Overall, in the initialization phase, the server distributes the FSL model to each client, usually containing the encoder and metric parameters. Each client locally trains with one epoch to obtain a local personalized bias features that matches the distribution of the respective data, which is then uploaded to the server to aggregate for a global bias feature to coordinate the subsequent training of the global and personalized parameters. It is worth noting that the global and local models do not have the same structure, the global model  $\phi$  has only encoder ( $\phi^f$ ) and metric ( $\phi^m$ ) parameters, while the local model  $\theta$  contains encoder ( $\theta^f$ ), classifier ( $\theta^c$ ), metric ( $\theta^m$ ), and reconstruction ( $\theta^r$ ) parameters, where only the metric parameters are shared. In the subsequent global communications, it is actually a two-stage learning process, starting with the decoupling of global and local features, where encoder and classifier are used. Then comes the realization of the decoupling of the category features, where encoder, metric and

---

**Algorithm 1: FedFSL-CFRD**


---

**Input:**  $M$ : the number of participating clients;  $\mathcal{D}_i$ : local data for each client  $c_i$ ;  $T$ : the number of global communication rounds;  $E_o$ : the number of epochs;  $E_i$ : the number of episodes;  $N$ : the number of classes in episodes;  $K$ : the number of labeled samples per category in the support set;  $Q$ : the number of unlabeled samples in the query set.

**Output:**  $\{\theta_1^*, \dots, \theta_M^*\}$ : personalized FSL parameters;  $\phi^*$ : global FSL parameter.

- 1: Initialize the local FSL parameters:  $\{\theta_1^0, \dots, \theta_M^0\} \leftarrow \phi^0$ .
  - 2: Initialize the local classifier parameter  $\{\theta_1^{c,0}, \dots, \theta_M^{c,0}\}$  and reconstruction parameter  $\{\theta_1^{r,0}, \dots, \theta_M^{r,0}\}$ .
  - 3: Train the local parameters  $\{\theta_i^{f,0}, \theta_i^{c,0}\}$  for each client  $c_i$  based on Eq. (1), and upload the local personalized bias features  $\{\bar{e}_1, \dots, \bar{e}_M\}$ .
  - 4: Server aggregates the global bias feature  $\bar{e} = \frac{1}{M} \sum_{i=1}^M \bar{e}_i$ .
  - 5: **for** global communication round  $t = 1$  to  $T$  **do**
  - 6:   Server sends the global parameter  $\phi^t$  to each client.
  - 7:   **for**  $i = 1, \dots, M$  in parallel **do**
  - 8:     Update the  $\phi_i^{f,t}, \theta_i^{f,t}, \theta_i^{c,t}$  based on Eq. (2) with  $E_{epo}$  epochs.
  - 9:     **for**  $\epsilon = 1$  to  $E_{epi}$  **do**
  - 10:        $\Gamma_\epsilon \leftarrow \sigma(\{1, \dots, C_{c_i}\}, N)$
  - 11:       **for**  $k$  in  $\Gamma_\epsilon$  **do**
  - 12:          $\mathcal{S}_k \leftarrow \sigma(\mathcal{D}_i^{\Gamma_k}, K)$ .
  - 13:          $\mathcal{Q}_k \leftarrow \sigma(\mathcal{D}_i^{\Gamma_k} \setminus \mathcal{S}_k, Q)$ .
  - 14:          $c_k = \frac{1}{N} \sum_{(\mathbf{x}_i, y_i) \in \mathcal{S}_k} (f_{\phi_i^{f,t}}(\mathbf{x}_i) + e_i^p)$ .
  - 15:       **end for**
  - 16:       Update the  $\theta_i^f, \theta_i^m$  and  $\theta_i^r$  based on Eqs. (4) and (5).
  - 17:     **end for**
  - 18:   **end for**
  - 19:   Server aggregates the global FSL parameter:  $\phi^t = \sum_{i=1}^M \frac{1}{M} \phi_i^t$ .
  - 20: **end for**
- 

reconstruction networks are used. The FedFSL-CFRD detailed process is given in Algorithm 1.

### Client-Wise Feature Disentanglement

In traditional FL (McMahan et al. 2017; Li et al. 2020), each client obtains the feature representation with forward propagation after receiving the global model, *i.e.*,  $e_i := f(\mathbf{x}_i, \phi_i)$ . However, due to the domain bias caused by heterogeneous data, the global parameter is not well personalized to appropriate the local data. More importantly, since FSL is based on the episodic training paradigm, it does not necessarily cover the local data distribution in each communication iteration. Therefore, we decouple the representations obtained by the feature extractor into global and personalized features, *i.e.*,  $e_i := e_i^g + e_i^p$ , where  $e_i^g$  and  $e_i^p$  are obtained by the encoder parameters in  $\phi_i^f$  and  $\theta_i^f$ , respectively.

In the initialization, after the clients receive the global parameters (usually containing an encoder  $\phi^f$  and a metric net-

work  $\phi^m$ ), the encoder and metric network of local parameters inherits the global parameters ( $\theta_i^0 \leftarrow \phi^0$ ), and initializes the classifier ( $\theta_i^{c,0}$ ) and discriminator ( $\theta_i^{d,0}$ ) parameters. Subsequently, the encoder and the classifier parameters are trained with  $\mathcal{D}_i$  in one epoch to obtain the personalized bias features  $\bar{e}_i := f(\mathcal{D}_i, \phi_i^f)$  as follows:

$$\mathcal{L}_{\mathcal{D}_i}(\theta_i^f, \theta_i^c) = \mathbb{E}_{(\mathbf{x}_i, y_i) \sim \mathcal{D}_i} [\ell_{CE}(h(f(\mathbf{x}_i; \theta_i^f); \theta_i^c), y_i)] \quad (1)$$

where  $\ell_{CE} = -\sum_{i=1}^{|\mathcal{D}_i|} (y_i \log \hat{y}_i)$  is the cross-entropy loss.

After initial training to obtain personalized bias features  $\{\bar{e}_1, \dots, \bar{e}_M\}$ , all the clients uploads them to the server for aggregation to obtain the global bias feature  $\bar{e} = \frac{1}{M} \sum_{i=1}^M \bar{e}_i$ . In the subsequent communication iterations,  $\bar{e}$  is frozen without further updates.

To maintain the privacy and personalized processing of client feature information, during the training process, we only update the personalized bias features locally without uploading it to the server. In the global communication, the global and local encoder parameters are decoupled to emphasize on general features and client-specific personalized features, respectively. The loss function is illustrated as:

$$\begin{aligned} \mathcal{L}_{\mathcal{D}_i}(\phi_i^f; \theta_i^f, \theta_i^c) = \\ \mathbb{E}_{(\mathbf{x}_i, y_i) \sim \mathcal{D}_i} [\ell(h(f(\mathbf{x}_i; \phi_i^f) + e_i^p; \theta_i^c), y_i)] + \zeta \|\bar{e}_i - \bar{e}\|^2 \end{aligned} \quad (2)$$

where  $e_i^p := f(\mathbf{x}_i; \theta_i^f)$ ,  $\zeta$  is a hyperparameter that controls the importance of global bias feature, while  $\bar{e}_i$  is updated by:

$$\bar{e}_i^t = (1 - \mu) \bar{e}_i^{t-1} + \mu \bar{e}_i^t \quad (3)$$

where  $\mu$  is a momentum hyperparameter that controls the weight given to the current batch,  $\bar{e}_i^{t-1}$  and  $\bar{e}_i^t$  are the average representation vectors of the previous and current batch, respectively.

### Class-Wise Feature Disentanglement

The client-wise feature disentanglement only decouples global and personalized representations for the encoder, *i.e.*, improves the discriminability of class-irrelevant representations. However, due to the data scarcity in FSL, the impact of class-relevance is crucial in only a small number of samples for the metric. Taking this cue, we design the class-wise feature disentanglement branch based on the few-shot branch, which improves the discriminability of class-relevance representations with self-reconfiguration and mutual-reconfiguration.

**FSL Branch:** We employ the episodic training paradigm to update the local parameters of FSL branch. It is first necessary to construct multiple episodes from  $\mathcal{D}_i$ . In the  $N$ -way  $K$ -shot setting,  $N$  categories are randomly selected from  $\mathcal{D}_i$ :  $\Gamma_\epsilon \leftarrow \sigma(\{1, \dots, C_{c_i}\}, N)$ , where  $\sigma$  stands for random sampling. Subsequently,  $K$  labeled samples are selected from each of these  $N$  categories to compose the support set ( $\mathcal{S}_k \leftarrow \sigma(\mathcal{D}_i^{\Gamma_k}, K)$ ), and  $Q$  samples are selected from the remaining samples in these  $N$  categories to form the query set ( $\mathcal{Q}_k \leftarrow \sigma(\mathcal{D}_i^{\Gamma_k} \setminus \mathcal{S}_k, Q)$ ). Then the prototypes for each category in the support set are obtained:

$c_k = \frac{1}{N} \sum_{(x_i, y_i) \in \mathcal{S}_k} (f_{\phi_i^{f,t}}(\mathbf{x}_i) + e_i^p)$ . Based on this, the parameters of the encoder and metric network are updated by the following loss function:

$$\begin{aligned} \mathcal{L}_{(x,y) \in \mathcal{Q}_k}(\phi_i^f; \theta_i^f, \theta_i^m) &= \frac{1}{N \times Q} [d((f_{\phi_i^f}(\mathbf{x}) + e_i^p), c_k) \\ &+ \log \sum_{k'} \exp(-d((f_{\phi_i^f}(\mathbf{x}) + e_i^p), c_{k'}))] \end{aligned} \quad (4)$$

where  $d$  is the metric function that measures the distance of the extracted representations, commonly known as the cosine distance in MatchingNet (Vinyals et al. 2016), Euclidean distance in ProtoNet (Snell, Swersky, and Zemel 2017) and relation module in Relation Net (Sung et al. 2018), etc. Here, we use  $\theta_i^m$  to implement a universal representation.

**Class-Relevance Reconstruction Branch:** Here, we achieve feature disentanglement fo class-relevance by constructing reconstruction and translation loss. To preserve the feature information and achieve the feature separation, we use a decoder  $D$  to combine for classification and FSL branches for feature self-reconstruction and class-reconstruction, where the decoder  $D$  is constructed from one fully connected block and three convolutional blocks. The specific loss function is as follows:

$$\begin{aligned} \mathcal{L}_{(x_i, y_i) \in \mathcal{S}_k \cup \mathcal{Q}_k}(\phi_i^f; \theta_i^f, \theta_i^r) &= \frac{1}{L} \sum_{i=1}^L \|f_{\phi_i^f}(\mathbf{x}) + e_i^p, \hat{\mathbf{x}}_i\|_1 \\ &+ \sum_{i=1}^L \|f_{\phi_i^f}(\mathbf{x}) + e_i^p, \hat{\mathbf{x}}_i^{y_i}\|_1 \end{aligned} \quad (5)$$

where  $\theta_i^r$  denotes the parameter of reconstruction network,  $L$  is the number of samples in  $\epsilon$ ,  $\|\cdot\|_1$  is the  $\ell_1$  norm.  $\hat{\mathbf{x}}_i$  and  $\hat{\mathbf{x}}_i^k$  are the reconstructed features by encoder  $D$  for sample  $x_i$  and prototype  $c_k$ .

Specifically, for each episode  $\epsilon$ , decoder  $D$  takes as input a pair of feature vectors from both branches, one is from an encoder consists of three convolutional blocks, the other is from the average pooling layer. It is aim to reconstruct the original features based on the source of class-related information for each sample  $x_i$ .

## Experiments

### Experimental Setup

**Datasets:** To demonstrate the effectiveness of our proposed FedFSL-CFRD, it is implemented in three benchmark few-shot datasets, *Mini-ImageNet* (Vinyals et al. 2016), *Tiered-ImageNet* (Ren et al. 2018) and *CIFAR-FS* (Bertinetto et al. 2018), respectively.

**FSL Homogeneous:** All the clients have the same categories in training and test sets, but the training and test categories do not intersect.

**FSL Heterogeneous:** Neither the training nor the testing categories are intersected between clients, and the respective training and testing samples are also non-overlapping.

**Implementation Details:** To ensure the fairness, all the experiments are implemented on dual-GPU NVIDIA GeForce RTX 4090, the overall classification accuracy is reported by the benchmark metrics 5-way 1-shot and 5-way 5-shot with 95% confidence intervals. The query samples in each meta-task is 15. we use the ProtoNet (Snell, Swersky, and Zemel 2017) with ResNet-12 backbone as the base model for all the FL comparison methods. The number of channels is [64, 160, 320, 640]. The number of clients  $M$  is set to 10 if nothing else is specified. For local model training, the initial value of the learning rate is set to 0.001, the SGD optimizer decay weight is 5e-4, and the momentum is 0.9. All comparison results are reported in the parameters realized for their best results. For the hyperparameters in FedFSL-CFRD, we set the  $E_o$  and  $E_i$  to 1 and 500, respectively.  $\zeta$  and  $\mu$  are set to 1 and 0.1, respectively.

### Comparison Results

To verify the superiority of our proposed FedFSL-CFRD, we chose 9 excellent peers for comparison, namely **FedAvg** (McMahan et al. 2017), **FedProx** (Li et al. 2020), **FedProto** (Tan et al. 2022b), **PerFedAvg** (Fallah, Mokhtari, and Ozdaglar 2020), **pFedMe** (T Dinh, Tran, and Nguyen 2020), **FedFomo** (Zhang et al. 2021b), **FedALA** (Zhang et al. 2023b), **GPFL** (Zhang et al. 2023a), and **FedFSL-Adv** (Fan and Huang 2021).

Table 1 gives the comparison results between FedFSL-CFRD and excellent peers in terms of 5-way 1-shot and 5-way 5-shot on three benchmark FSL datasets. The experimental results show that FedFSL-CFRD has a significant advantage both in relatively simple and complex scenarios. Traditional FL algorithms (such as FedAvg, FedProx) can achieve decent performance, but there is still a gap compared to PFL algorithms, especially under data heterogeneity. For PFL algorithms, most of them are not designed for FSL scenarios, and thus will perform slightly less well than FedFSL methods (FedFSL-Adv, FedFSL-CFRD) in generalizing to novel classes. In contrast to FedFSL-ADV, which is formulated adversarially to train and optimize the local models to produce samples that can better represent unseen data samples, it is equivalent to adding additional training samples. While FedFSL-CFRD only utilizes the current limited data for decoupling design in feature representation, and outperforms FedFSL-ADV which requires additional pseudo-samples in overall performance by about 1% in both 5-way 1-shot and 5-way 5-shot on three benchmark FSL datasets.

Besides, in addition to the generalizability comparison of the global model, we also present the convergence of the client personalized models as shown in Figure 3. It can be concluded that FedFSL-CFRD achieves stable convergence of the model in fewer communication rounds, which is largely due to the fact that by decoupling the collaborative features of the client-wise and the class-wise, it can improve the discriminative ability for the local data representations to a certain extent, and thus improve the performance of the metrics in the scarcity of samples. Therefore, FedFSL-CFRD is not only superior in few-shot accuracy, but also capable of being well in convergence.

Method	<i>Mini-ImageNet</i>		<i>Tiered-ImageNet</i>		<i>CIFAR-FS</i>	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
<i>FSL Homogeneous</i>						
FedAvg	51.32±1.41	66.67±2.01	55.63±1.27	73.72±0.67	58.15±0.92	71.05±1.51
FedProx	51.71±1.92	67.16±1.50	56.27±1.31	74.02±1.42	58.90±1.64	71.67±0.97
FedProto	52.39±1.34	67.77±2.07	56.91±0.91	75.35±0.97	59.66±1.33	72.14±1.36
PerFedAvg	53.09±1.36	68.40±1.26	57.50±1.78	75.39±1.17	60.40±1.79	72.72±1.00
pFedMe	53.45±1.56	69.06±1.29	57.99±1.18	76.59±1.62	60.87±1.23	73.59±1.39
FedFomo	54.13±1.97	69.66±1.59	58.65±1.44	77.52±1.27	61.38±1.50	73.39±1.60
FedALA	56.05±1.53	71.36±1.25	60.03±1.05	78.05±1.36	61.97±1.85	74.06±1.16
GPFL	56.10±1.74	71.41±1.14	60.14±1.55	78.56±1.48	62.58±1.84	74.17±1.12
FedFSL-Adv	55.75±1.65	70.59±1.64	59.88±1.38	78.19±1.38	62.45±1.75	74.02±1.84
<b>FedFSL-CFRD</b>	<b>56.66±1.52</b>	<b>72.15±0.98</b>	<b>60.52±1.22</b>	<b>78.95±1.28</b>	<b>63.00±1.16</b>	<b>74.58±0.89</b>
<i>FSL Heterogeneous</i>						
FedAvg	50.05±1.68	65.09±1.90	53.88±1.41	71.25±1.61	56.15±0.84	69.96±1.20
FedProx	50.76±1.09	65.70±0.80	54.26±1.42	71.96±1.40	56.74±1.57	69.99±0.97
FedProto	51.30±1.50	66.16±1.57	54.51±1.63	72.41±1.23	57.28±1.30	70.42±1.25
PerFedAvg	51.72±1.32	66.68±0.82	55.10±1.26	73.17±1.45	57.51±1.72	70.02±1.98
pFedMe	52.04±1.65	67.11±1.03	55.49±1.51	73.59±1.39	57.60±0.88	71.29±1.69
FedFomo	52.47±1.35	67.77±1.08	55.72±1.56	74.37±1.60	57.83±1.66	70.79±0.79
FedALA	52.96±0.96	69.64±1.44	57.25±1.46	75.86±1.70	58.96±1.21	71.17±1.55
GPFL	53.15±1.52	69.53±1.36	57.47±1.45	76.05±1.03	59.21±0.83	71.95±1.52
FedFSL-Adv	52.52±1.68	69.56±1.86	57.34±1.70	75.74±1.06	58.93±1.18	72.00±1.80
<b>FedFSL-CFRD</b>	<b>53.89±1.06</b>	<b>70.42±1.33</b>	<b>58.04±1.00</b>	<b>76.64±1.24</b>	<b>59.36±1.02</b>	<b>72.44±0.96</b>

Table 1: The test accuracy (in %) in 5-way 1-shot and 5-way 5-shot on three benchmark FSL datasets.

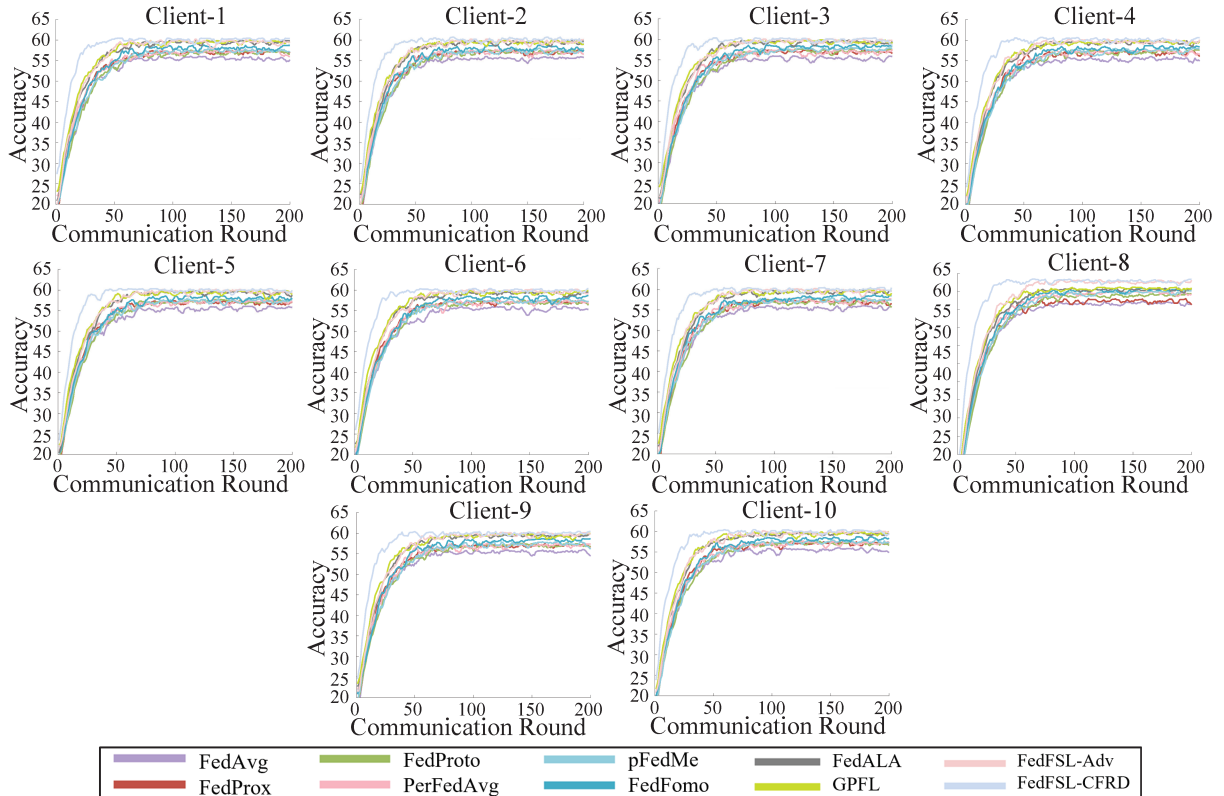


Figure 3: The test comparative accuracy (in %) of the local personalized models in the case of 5-way 1-shot (in FSL Homogeneous) on the Tiered-ImageNet.

Method	Mini-ImageNet		Tiered-ImageNet	
	1-shot	5-shot	1-shot	5-shot
	<i>FSL Homogeneous</i>			
Basic	51.32±1.41	66.67±2.01	55.63±1.27	73.72±0.67
Basic+Client-Wise Feature Disentanglement	51.96±1.75	67.02±1.94	55.97±1.45	73.96±1.53
Basic+Class-Wise Feature Disentanglement	53.29±1.68	69.73±1.64	58.06±1.39	76.65±1.64
<b>FedFSL-CFRD</b>	<b>56.66±1.52</b>	<b>72.15±0.98</b>	<b>60.52±1.22</b>	<b>78.95±1.28</b>
	<i>FSL Heterogeneous</i>			
Basic	50.05±1.68	65.09±1.90	53.88±1.41	71.25±1.61
Basic+Client-Wise Feature Disentanglement	50.83±1.70	66.26±1.83	54.46±1.52	72.84±1.75
Basic+Class-Wise Feature Disentanglement	51.48±1.52	68.85±1.79	56.75±1.46	75.91±1.54
<b>FedFSL-CFRD</b>	<b>53.89±1.06</b>	<b>70.42±1.33</b>	<b>58.04±1.00</b>	<b>76.64±1.24</b>

Table 2: The ablation results (in %) in 5-way 1-shot and 5-way 5-shot on Mini-ImageNet and Tiered-ImageNet datasets.

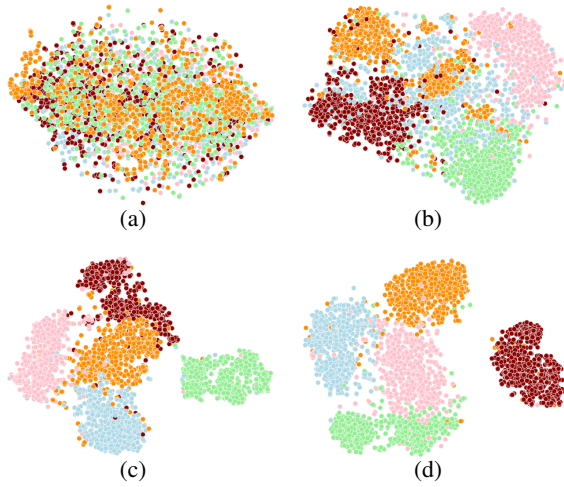


Figure 4: t-SNE visualization comparison results on the Mini-ImageNet dataset. (a) Basic; (b) Basic+Client-Wise Feature Disentanglement; (c) Basic+Class-Wise Feature Disentanglement; (d) FedFSL-CFRD.

### Ablation Study

In ablation experiments, we validate the effectiveness of the individual components of FedFSL-CFRD on Mini-ImageNet and Tiered-ImageNet datasets. Table 2 presents the ablation experiments for designing client-wise and Class-wise feature disentanglement on the FedFSL framework, where “Basic” refers to the global global aggregation based on the ProtoNet (Snell, Swersky, and Zemel 2017) according to the FSL paradigm.

To make the results more intuitive, Figures 4 and 5 give the t-SNE visualization results of FedFSL-CFRD in both components and all combined in Mini-ImageNet and Tiered-ImageNet, respectively. It is straightforward to observe from the figure that when there is no feature disentanglement, the basic model does not have discriminative ability for complex features. When client-wise feature disentanglement is designed, the encoder is able to discriminate the categories to a certain extent due to the decoupling of the personalization parameters, but it is limited to the discrimination of

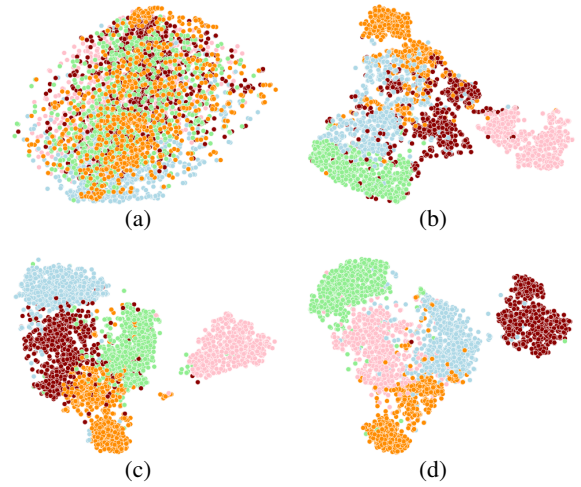


Figure 5: t-SNE visualization comparison results on the Tiered-ImageNet dataset. (a) Basic; (b) Basic+Client-Wise Feature Disentanglement; (c) Basic+Class-Wise Feature Disentanglement; (d) FedFSL-CFRD.

class-irrelevance. By adding the class-wise feature disentanglement, the encoder is able to enhance the discrimination of class-relevance in FedFSL-CFRD.

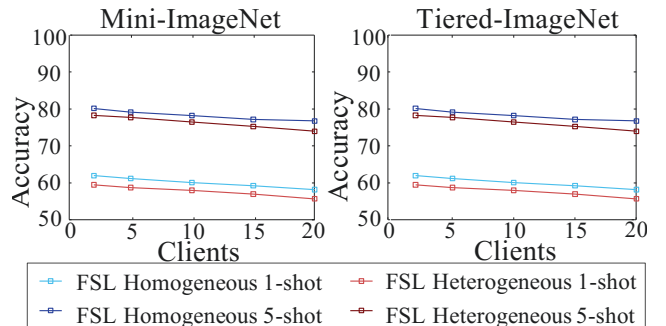


Figure 6: The effect of the number of participating clients in FedFSL-CFRD.

Moreover, Figure 6 provides the effect of the number of

participating clients in FedFSL-CFRD. It is worth noting that the total volume of data is constant, and an increase in the number of clients reduces the amount of data or the number of categories each client has, implying an exacerbation of the heterogeneity problem in the FL paradigm. Therefore, the performance degradation of homogeneous data is less obvious than that of heterogeneous data. As the number of clients increases, the performance of FedFSL-CFRD decreases slightly but not significantly, due to the fact that our client-wise feature disentanglement in global generalized features and local personalized feature can mitigate the alleviate the effects of data heterogeneity to a certain extent.

## Conclusion

In this work, we presented a novel collaborative feature representation disentanglement (CFRD) method tailored for federated few-shot learning (FedFSL) scenarios. Our approach addresses the key challenges of data heterogeneity across clients and the scarcity of labeled data within individual clients by decoupling features at both the client-wise and class-wise. By distinguishing between global aggregation parameter and local personalized parameter, as well as class-specific and class-agnostic features, CFRD effectively enhances the discrimination ability to generalize across diverse data distributions and quickly adapt to novel classes. Extensive experiments confirmed the effectiveness of FedFSL-CFRD.

## Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant Nos. 62476210, 62036006), Natural Science Basic Research Program of Shaanxi (Program Nos. 2024JC-YBMS-485, 2024JC-YBQN-0633).

## References

- Bertinetto, L.; Henriques, J.; Torr, P.; and Vedaldi, A. 2018. Meta-learning with differentiable closed-form solvers. In *ICLR*.
- Chen, J.; and Zhang, A. 2024. On Disentanglement of Asymmetrical Knowledge Transfer for Modality-Task Agnostic Federated Learning. In *AAAI*, volume 38, 11311–11319.
- Chen, Y.; Liu, Z.; Xu, H.; Darrell, T.; and Wang, X. 2021. Meta-baseline: Exploring simple meta-learning for few-shot learning. In *ICCV*, 9062–9071.
- Cheng, H.; Wang, Y.; Li, H.; Kot, A. C.; and Wen, B. 2023. Disentangled feature representation for few-shot image classification. *IEEE TNNLS*.
- Collins, L.; Hassani, H.; Mokhtari, A.; and Shakkottai, S. 2021. Exploiting shared representations for personalized federated learning. In *ICML*, 2089–2099.
- Dang, Z.; Luo, M.; Wang, J.; Jia, C.; Yan, C.; Dai, G.; Chang, X.; and Zheng, Q. 2024. Disentangled Generation with Information Bottleneck for Enhanced Few-Shot Learning. *IEEE TIP*.
- Duan, M.; Liu, D.; Chen, X.; Liu, R.; Tan, Y.; and Liang, L. 2020. Self-balancing federated learning with global imbalanced data in mobile systems. *IEEE TPDS*, 32(1): 59–71.
- Fallah, A.; Mokhtari, A.; and Ozdaglar, A. 2020. Personalized federated learning: A meta-learning approach. *arXiv preprint arXiv:2002.07948*.
- Fan, C.; and Huang, J. 2021. Federated few-shot learning with adversarial learning. In *2021 19th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 1–8. IEEE.
- Huang, W.; Ye, M.; Shi, Z.; Li, H.; and Du, B. 2023. Rethinking federated learning with domain shift: A prototype view. In *CVPR*, 16312–16322.
- Jamal, M. A.; and Qi, G.-J. 2019. Task agnostic meta-learning for few-shot learning. In *CVPR*, 11719–11727.
- Jeong, E.; Oh, S.; Kim, H.; Park, J.; Bennis, M.; and Kim, S.-L. 2018. Communication-efficient on-device machine learning: Federated distillation and augmentation under non-iid private data. *arXiv preprint arXiv:1811.11479*.
- Laenen, S.; and Bertinetto, L. 2021. On episodes, prototypical networks, and few-shot learning. In *NerulIPS*, volume 34, 24581–24592.
- Li, J.; Zhang, Y.; Qiang, W.; Si, L.; Jiao, C.; Hu, X.; Zheng, C.; and Sun, F. 2023. Disentangle and remerge: interventional knowledge distillation for few-shot object detection from a conditional causal perspective. In *AAAI*, volume 37, 1323–1333.
- Li, Q.; He, B.; and Song, D. 2021. Model-contrastive federated learning. In *CVPR*, 10713–10722.
- Li, T.; Hu, S.; Beirami, A.; and Smith, V. 2021. Ditto: Fair and robust federated learning through personalization. In *ICML*, 6357–6368.
- Li, T.; Sahu, A. K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; and Smith, V. 2020. Federated optimization in heterogeneous networks. In *MLSys*, volume 2, 429–450.
- Lin, C.-C.; Chu, H.-L.; Wang, Y.-C. F.; and Lei, C.-L. 2021. Joint feature disentanglement and hallucination for few-shot image classification. *IEEE TIP*, 30: 9245–9258.
- Lin, T.; Kong, L.; Stich, S. U.; and Jaggi, M. 2020. Ensemble distillation for robust model fusion in federated learning. In *NerulIPS*, volume 33, 2351–2363.
- McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial Intelligence and Statistics*, 1273–1282.
- Mohri, M.; Sivek, G.; and Suresh, A. T. 2019. Agnostic federated learning. In *ICML*, 4615–4625.
- Qiao, L.; Shi, Y.; Li, J.; Wang, Y.; Huang, T.; and Tian, Y. 2019. Transductive episodic-wise adaptive metric for few-shot learning. In *ICCV*, 3603–3612.
- Ren, M.; Triantafillou, E.; Ravi, S.; Snell, J.; Swersky, K.; Tenenbaum, J. B.; Larochelle, H.; and Zemel, R. S. 2018. Meta-Learning for Semi-Supervised Few-Shot Classification. In *ICLR*.



- Shome, D.; and Kar, T. 2021. FedAffect: Few-shot federated learning for facial expression recognition. In *ICCV*, 4168–4175.
- Snell, J.; Swersky, K.; and Zemel, R. 2017. Prototypical networks for few-shot learning. In *NerulPS*, volume 30.
- Sun, B.; Huo, H.; Yang, Y.; and Bai, B. 2021. Partialfed: Cross-domain personalized federated learning via partial initialization. In *NerulPS*, volume 34, 23309–23320.
- Sun, J.; Chen, T.; Giannakis, G. B.; Yang, Q.; and Yang, Z. 2020. Lazily aggregated quantized gradient innovation for communication-efficient federated learning. *IEEE TPAMI*, 44(4): 2031–2044.
- Sun, Q.; Liu, Y.; Chua, T.-S.; and Schiele, B. 2019. Meta-transfer learning for few-shot learning. In *CVPR*, 403–412.
- Sun, X.; Yang, S.; and Zhao, C. 2022. Lightweight industrial image classifier based on federated few-shot learning. *IEEE TII*, 19(6): 7367–7376.
- Sung, F.; Yang, Y.; Zhang, L.; Xiang, T.; Torr, P. H.; and Hospedales, T. M. 2018. Learning to compare: Relation network for few-shot learning. In *CVPR*, 1199–1208.
- T Dinh, C.; Tran, N.; and Nguyen, J. 2020. Personalized federated learning with moreau envelopes. In *NerulPS*, volume 33, 21394–21405.
- Tan, A. Z.; Yu, H.; Cui, L.; and Yang, Q. 2022a. Towards personalized federated learning. *IEEE TNNLS*, 34(12): 9587–9603.
- Tan, Y.; Long, G.; Liu, L.; Zhou, T.; Lu, Q.; Jiang, J.; and Zhang, C. 2022b. Fedproto: Federated prototype learning across heterogeneous clients. In *AAAI*, volume 36, 8432–8440.
- Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D.; et al. 2016. Matching networks for one shot learning. In *NerulPS*, volume 29.
- Wang, J.; Liu, Q.; Liang, H.; Joshi, G.; and Poor, H. V. 2020. Tackling the objective inconsistency problem in heterogeneous federated optimization. In *NerulPS*, volume 33, 7611–7623.
- Wang, J.; Yang, X.; Cui, S.; Che, L.; Lyu, L.; Xu, D. D.; and Ma, F. 2024. Towards personalized federated learning via heterogeneous model reassembly. In *NerulPS*, volume 36.
- Yang, C.; Zhu, M.; Liu, Y.; and Yuan, Y. 2023. FedPD: Federated Open Set Recognition with Parameter Disentanglement. In *ICCV*, 4882–4891.
- Yu, Z.; Chen, L.; Cheng, Z.; and Luo, J. 2020. Transmatch: A transfer-learning scheme for semi-supervised few-shot learning. In *CVPR*, 12856–12864.
- Zhang, J.; Hua, Y.; Wang, H.; Song, T.; Xue, Z.; Ma, R.; Cao, J.; and Guan, H. 2023a. Gpfl: Simultaneously learning global and personalized feature information for personalized federated learning. In *ICCV*, 5041–5051.
- Zhang, J.; Hua, Y.; Wang, H.; Song, T.; Xue, Z.; Ma, R.; and Guan, H. 2023b. Fedala: Adaptive local aggregation for personalized federated learning. In *AAAI*, volume 37, 11237–11244.
- Zhang, J.; Wang, Y.; Zhu, K.; Zhang, Y.; and Li, Y. 2021a. Diagnosis of interturn short-circuit faults in permanent magnet synchronous motors based on few-shot learning under a federated learning framework. *IEEE TII*, 17(12): 8495–8504.
- Zhang, M.; Sapra, K.; Fidler, S.; Yeung, S.; and Alvarez, J. M. 2021b. Personalized Federated Learning with First Order Model Optimization. In *ICLR*.
- Zhao, Y.; Yu, G.; Wang, J.; Domeniconi, C.; Guo, M.; Zhang, X.; and Cui, L. 2022. Personalized federated few-shot learning. *IEEE TNNLS*, 35(2): 2534–2544.
- Zhu, Z.; Hong, J.; and Zhou, J. 2021. Data-free knowledge distillation for heterogeneous federated learning. In *ICML*, 12878–12889.