Cross-Silo Feature Space Alignment for Federated Learning on Clients with Imbalanced Data

Zhuang Qi[1], Lei Meng[1,2*], Zhaochuan Li[3], Han Hu[4], Xiangxu Meng[1]

[1]School of Software, Shandong University, China
[2]Shandong Research Institute of Industrial Technology, China
[3]Inspur, China
[4]School of information and Electronics, Beijing Institute of Technology, China
z_qi@mail.sdu.edu.cn, lmeng@sdu.edu.cn, lizhaoch@inspur.com, hhu@bit.edu.cn, mxx@sdu.edu.cn

Abstract

Data imbalance across clients in federated learning often leads to different local feature space partitions, harming the global model's generalization ability. Existing methods either employ knowledge distillation to guide consistent local training or performs procedures to calibrate local models before aggregation. However, they overlook the ill-posed model aggregation caused by imbalanced representation learning. To address this issue, this paper presents a cross-silo feature space alignment method (FedFSA), which learns a unified feature space for clients to bridge inconsistency. Specifically, FedFSA consists of two modules, where the in-silo prototypical space learning (ISPSL) module uses predefined text embeddings to regularize representation learning, which can improve the distinguishability of representations on imbalanced data. Subsequently, it introduces a variance transfer approach to construct the prototypical space, which aids in calibrating minority classes feature distribution and provides necessary information for the cross-silo feature space alignment (CSFSA) module. Moreover, the CSFSA module utilizes augmented features learned from the ISPSL module to learn a generalized mapping and align these features from different sources into a common space, which mitigates the negative impact caused by imbalanced factors. Experimental results from three datasets verified that FedFSA improves the consistency between diverse spaces on imbalanced data, which results in superior performance compared to existing methods. The source codes have been released at https://github.com/qizhuang-qz/FedFSA.

Introduction

Federated learning enables collaborative modeling with imbalanced data from various sources, which shares model parameters instead of raw data between data sources and the server (Hu et al. 2024b; Liu et al. 2023; Cai et al. 2024b; Qi et al. 2022; Kairouz et al. 2021). This significantly improves the effective utilization of isolated data, enabling them to contribute to cooperative decision-making and learn a generalized model (Cai et al. 2024a; Wang et al. 2023a; Meng et al. 2024; Wang et al. 2024a). However, existing studies show that the data heterogeneity between clients could lead to a decrease in the effectiveness of collaborative modeling (Qi et al. 2024a; Shi et al. 2023; Wen et al. 2023; Qi et al. 2024b). This is mainly because learning a consistent feature space becomes challenging when dealing with data that is imbalanced within clients and has inconsistent distribution across clients, which makes it difficult to integrate learners with inconsistent objectives into a remarkable model.

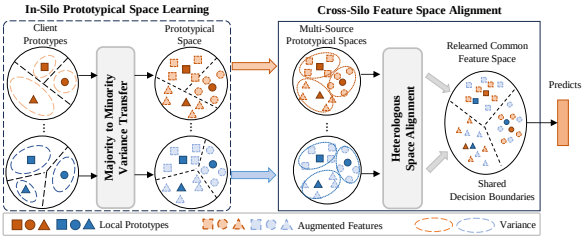


Figure 1: FedFSA leverages the variance transfer approach at the client side to learn the prototypical space, which calibrates the feature distribution of minority classes. Moreover, FedFSA aligns feature spaces from different sources at the server side to learn shared decision boundaries.

To alleviate issues of class imbalance, existing methods can roughly be categorized into two types. The former approach typically employs knowledge distillation to guide local model learning on the client-side, which aims to transfer global knowledge to client models and leverage regularization techniques to guide them in learning consistent representations of data (MOON (Li, He, and Song 2021), Fedproc (Mu et al. 2023), FedNTD (Lee et al. 2022) and FPL (Huang et al. 2023)). For instance, Fedproc and FPL construct prototypical representation for each class of samples and employ it to facilitate the feature space alignment across clients. The latter method usually involves model calibration, including global classifier fine-tuning and projection head retraining, which is aimed at mitigating bias issues introduced by the weighted averaging of local models (CReFF (Shang et al. 2022), CLIP2FL (Shi et al. 2024) and FedCSPC (Qi et al. 2023)), where CReFF and CLIP2FL focus on refining the global classifier to enhance its robustness with varied feature environments. Notably, these strategies have shown promising results in classes with a majority of samples. However, data imbalance typically results in poor representation learning for minority classes in clients. Additionally, they directly

*Corresponding author

use the intermediate features output by clients to retrain the model, but the inherent differences between features from different sources limit their effectiveness.

To address this problem, this paper presents a **c**ross-**s**ilo **f**eature **s**pace **a**lignment method, termed FedFSA, which constructs a unified space to align features from diverse sources to mitigate the negative impact of data imbalance. As depicted in Figure 1, FedFSA includes two main modules, including the in-silo prototypical space learning (ISPSL) module and the cross-silo feature space alignment (CSFSA) module. Specifically, the ISPSL module leverages pre-defined feature sapce learned from pretrained CLIP to guide the local representation learning, which is conducive to enhance the discriminability of the minority class representation in imbalanced data. Moreover, the ISPSL module introduces the variance transfer technique that leverage the diversity of samples in majority classes to expand and construct the prototypical space of minority classes. This provides meaningful and privacy-preserving information for the cross-silo feature space alignment (CSFSA) module. Subsequently, the CSFSA module maps features from diverse spaces into a unified space to further reduce feature distribution discrepancies between data sources caused by class imbalanced and emphasizes the contribution of each feature by weighting them based on their attention scores, effectively mitigating the interference of outliers.

Experiments were conducted on three datasets, including performance comparison, ablation study, in-depth analysis, case study and error analysis of FedFSA. The results validate that FedFSA can promote precise cross-silo feature alignment on imbalanced data. Furthermore, the error analysis can offer valuable insights to guide future refinements. In summary, the main contributions of this paper include:

- To alleviate the negative impact of data imbalance in federated learning, this study proposes a cross-silo feature space alignment method (FedFSA). To the best of our knowledge, FedFSA is the first method to align feature spaces from different sources on imbalanced data.
- This study proposes a model-agnostic framework, which can integrate various client-based methods. It mitigates the impact of imbalanced data by learning a shared feature space for different clients.
- Experimental findings have revealed that aligning the feature spaces of different clients can benefit the retrained model, which avoids the impact of inherent differences between client feature spaces. This provides a feasible approach for future research.

Related Work

Methods based on knowledge distillation

Knowledge distillation methods aim to guide clients to learn consistent knowledge, which mitigates data imbalance in federated learning. Typically, they entail the use of extra information as a regularizer to regulate updates locally (Li, He, and Song 2021; Tan et al. 2022; Huang et al. 2023; Yu et al. 2021; Ye et al. 2023; Li et al. 2024b; Ren et al. 2024). Within the context, regularization have played a significant role (Wu et al. 2023). For example, MOON employs contrastive regularization to penalize inconsistencies between local and global feature spaces (Li, He, and Song 2021). FedProc (Mu et al. 2023), FedProto (Tan et al. 2022) and FPL (Huang et al. 2023) construct prototypes for each class based on data representations to represent the center of within-class representations. It guides the local training process by constraining the representations of all clients to converge towards these prototypes. Moreover, guiding the calibration of the feature space with a classifier is also an effective strategy, such as FedETF employs a fixed simplex equiangular tight frame classifier to encourage all clients in learning a unified and optimal feature representation (Li et al. 2023). AdressIM infers the global data distribution and mitigates global imbalance by using a ratio-weighted approach (Wang et al. 2021). Despite the positive outcomes of these methods, further exploration is needed for imbalanced data, as imbalances typically accumulate errors across training iterations.

Methods based on model calibration

Different from knowledge distillation, model calibration methods concentrate on making improvements on the server side, which re-trains global model to alleviate class imbalance issues. These methods along this line including global classifier calibration (Luo et al. 2021; Shang et al. 2022; Zeng et al. 2023; Shi et al. 2024), projection head retraining (Qi et al. 2023), and global model fine-tuning (Zhang et al. 2022; Hu et al. 2024a,c). They both hope to obtain a generalized model to fit all data from various sources. For instance, CCVR fuses the mean and variance of sample features obtained from client and employs a gaussian model to generate virtual features for retraining the global classifier (Luo et al. 2021). CReFF (Shang et al. 2022) and CLIP2FL (Shi et al. 2024) generate a series of federated features with gradients consistent with real data to fine-tune the classifier. FedFTG transfers knowledge from local to global models by exploring input spaces with a generator to fine-tune the entire global model (Zhang et al. 2022). From these analysis, their performance is closely tied to the quality of local feature information. However, both of them overlook the interference of local imbalance.

Problem Formulation

Federated learning systems typically utilize multiple data sources to collaboratively build the global model. It contains K data sources, $\mathcal{S} = \{S_1, ..., S_K\}$, and a central sever S. The source S_k utilizes its private data $D^k = \{(X^k, Y^k)\}$ to optimize the model M_k with the objective $\ell_k(\theta_k; D^k)$, where θ_k is the parameter of the model M_k. And the server S aggregates the parameters of all locally learned models $\{\theta_k | k = 1, ..., K\}$ to obtain global parameters, i.e., $\theta_g = \sum_{k=1}^{K} \alpha_k \theta_k$, where where $\alpha_k = |D^k| / \bigcup_{k=1}^{K} |D^k|$.

By comparison, FedFSA introduces the in-silo feature space reconstruction (ISPSL) module and the cross-silo feature space alignment (CSFSA) module, where the ISPSL module improves representation learning by aligning image representations f_i with label text embeddings $U = \{u_1, ..., u_C\}$ (where C is the number of classes) learned

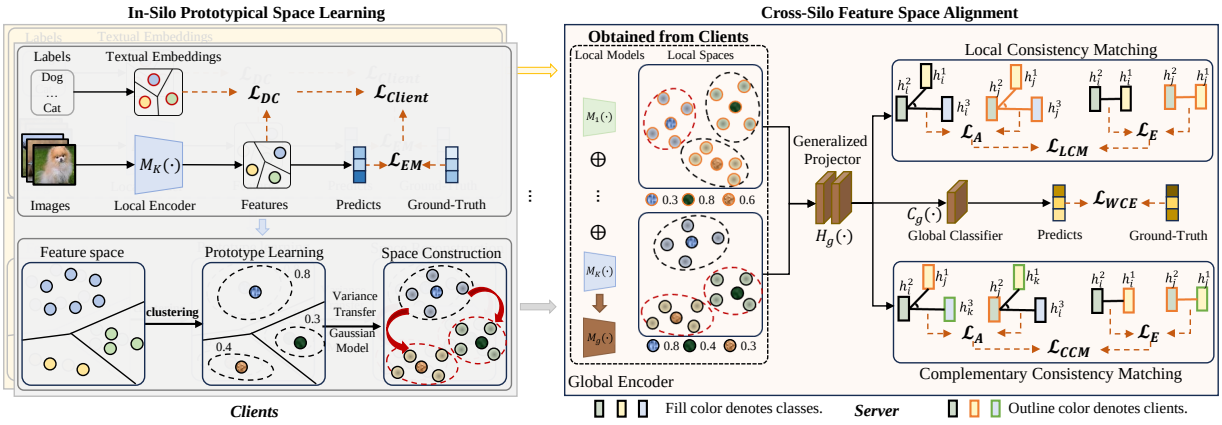


Figure 2: The in-silo prototypical space learning module utilizes label textual embeddings learned from pretrained CLIP to regularize the feature learning. Moreover, it transfers variance from the majority to the minority class to construct the prototypical space. Finally, the cross-silo feature space alignment module aligns feature spaces from different data sources.

from the pre-trained CLIP model, i.e., $f_i \mapsto u_i$. Meanwhile, the ISPSL module uses clustering to learn the cluster prototype μ_k^i, cluster variance Σ_k^i, and attention score s_k^i. Furthermore, the ISPSL module generates augmented features f_a based on the variance Σ ($\mu_k^i \oplus \mathcal{N}(0, \Sigma) \mapsto f_a$), where $\mathcal{N}(\cdot)$ is a Gaussian model. Subsequently, the CSFSA module leverages these augmented features to relearn the generalized projection $H_g(\cdot)$ and classifier $C_g(\cdot)$ on the server, i.e., $H_g(\mu_k^i) \approx H_g(\mu_K^i)$. And, FedFSA optimizes $H_g(\cdot)$ and $C_g(\cdot)$ based on alignment and classification status.

Approach

This study proposes a cross-silo feature space alignment method (FedFSA) in federated learning, which alleviates the issue of ill-posed aggregation caused by imbalanced data across clients. As shown in Figure 2, FedFSA includes two modules: the in-silo prototypical space learning (ISPSL) module and the cross-silo feature space alignment (CSFSA) module. Specifically, the ISPSL module transfers variance knowledge from majority to minority class to calibrate feature distribution and provides feature information to the CSFSA module while preserving privacy. The CSFSA module aligns feature spaces from different sources to bridge feature gaps between clients, which forms a generalized model.

In-Silo Prototypical Space Learning (ISPSL)

The ISPSL module aims to enhance representation learning and construct prototypical spaces on imbalanced data, which mitigates the impact of imbalances and provides image features for the CSFSA module, while preserving data privacy. However, data imbalance leads to a decline in the discriminability of minority class features, which compromises the effectiveness of cross-silo feature space alignment. To address this issue, the ISPSL module designs two processes: text-enhanced representation learning and variance transfer based space construction.

Text-Enhanced Representation Learning (TERL). To enhance the discriminability of minority class features, the TERL module uses a predefined space from pre-trained CLIP (Radford et al. 2021) to regularize representation learning. Specifically, it employs supervised prototypical contrastive learning to align image feature f_k^c with textual embedding u_c in client k. The loss $\mathcal{L}_{DC}$ is defined as:

$$\mathcal{L}_{DC} = -\frac{1}{N_k}\sum_{i=1}^{N_k}\log\frac{\sum_{c=1}^{C}\mathbb{1}_{y_k^c=c}\exp(f_k^c \cdot u_c/\tau)}{\sum_{c=1}^{C}\exp(f_k^c \cdot u_c/\tau)}, \quad (1)$$

where $u_c = CLIP_{\text{text}}$('a photo of [the name of class c]'), $CLIP_{\text{text}}(\cdot)$ is the text encoder. C denotes the quantity of classes. y_k^c is the label of f_k^c, $\mathbb{1}_{\text{True}} = 1$ and $\mathbb{1}_{\text{False}} = 0$, N_k is the number of training data of client k. Meanwhile, the TERL module uses an empirical loss to ensure the discriminative capability of the model, i.e.,

$$\mathcal{L}_{EM} = -\frac{1}{N_k}\sum_{i=1}^{N_k}\left(\sum_{c=1}^{C} y_c z_c + \sum_{c=1}^{C} y_c \log\left(\sum_{j=1}^{C} e^{z_j}\right)\right), \quad (2)$$

where z_c represents the c-th element in the model output vector. y_c is the ground-truth of image.

Variance Transfer based Space Construction (VTSC). To provide features for the CSFSA module in a privacy-preserving manner, the VTSC module constructs the prototypical space. It sends augmented features to the server instead of the original features, which distinguishes it from existing methods (Chen et al. 2024; Yang et al. 2024b). Despite efforts to improve feature learning, intra-class variations and inter-class overlap create noise that disrupts prototype modeling. Therefore, the VTSC module first employs clustering (Meng, Tan, and Miao 2019; Meng, Tan, and Wunsch 2015; Qi et al. 2023) to mine patterns in the latent space within each class and evaluates the importance of each prototype,

$$v_j^1, ..., v_j^{N_v} = \text{clustering}(M_k(D_j^k), N_v), \quad (3)$$

where $M_k(\cdot)$ is local model with frozen parameters and D_j^k denotes data of class j in client k. N_v is the number of clusters. Subsequently, the VSTR module calculates the mean $\mu_j^{N_v}$ and variance $\Sigma_j^{N_v}$ of features within clusters $v_j^{N_v}$, i.e.,

$$\mu_j^{N_v} = \frac{1}{N}\sum_{i=1}^{|v_j^{N_v}|} f_i \in v_j^{N_v}, \quad (4)$$

$$\Sigma_j^{N_v} = \frac{1}{|v_j^{N_v}|-1}\sum_{i=1}^{|v_j^{N_v}|}\left(f_i - \mu_j^{N_v}\right)\left(f_i - \mu_j^{N_v}\right)^T, \quad (5)$$

where $|\cdot|$ denotes the size of cluster. To reduce interference of outlier in model calibration, the VSTR module evaluates cluster significance using three factors: cluster size (ρ), compactness (σ), and minimum distance to other cluster centers (ξ). For cluster v_j^t, $\rho_j^t = |v_j^t|$, $\sigma_j^t = \frac{1}{n}\sum_{i=1}^{|v_j^t|}\left\|f_i - \mu_j^t\right\|_2$, $\xi_j^t = min\{\left\|\mu - \mu_j^t\right\|_2\}$, where μ is the cluster center of a different class than v_j^t, f_i is a feature of cluster v_j^t. Inherently, the larger, more compact clusters farther from others are more important. Therefore, the importance score of cluster v_j^t is $s_j^t = \frac{\rho_j^t \times \xi_j^t}{\sigma_j^t}$.

Moreover, data imbalance prevents minority class samples from adequately covering the underlying distribution, which is shown in long-tailed learning (Li et al. 2024a, 2021). Consequently, the VTSC module transfer variance from majority to minority classes, which calibrates the feature distribution on imbalanced data. Specifically, the VTSC module uses a Gaussian model $\mathcal{N}(\cdot)$ to generate augmented features based on variance (Lindsay 1995; Nie et al. 2016). It fuses the variance Σ_{maj} of a randomly selected majority class and other variances to transfer distribution knowledge to other features μ in a client, i.e.,

$$\left\{f_a^j\right\} = \{\mu + \Delta_j \mid \Delta_j \in \mathcal{N}\left(0, \Sigma_{fuse}\right), j = 1, ..., J\}, \quad (6)$$

where μ is a local prototype. J is the number of augmented features $\{f_a^j\}$, $\Sigma_{fuse} = (1-\kappa)*\Sigma + \kappa*\Sigma_{maj}$ denotes fused variance. These augmented features share the same scores as their corresponding prototypes. And the VTSC module transmits all augmented features, their corresponding scores, and the local model from the client to the CSFSA module.

Cross-Silo Feature Space Alignment (CSFSA)

The CSFSA module aims to learn a generalized model that fits the data from all clients. However, the inherent feature differences between different sources severely limit the performance of the retrained model. To address this problem, the CSFSA module employs cross-silo feature space alignment to map features from different sources into a unified space, which is used to bridge inconsistency between clients caused by data imbalance.

Specifically, the CSFSA module uses features reconstructed from the VTSC module to learn a generalized projection $H_g(\cdot)$ and classifier $C_g(\cdot)$, which enables global model to realize the unified feature learning for samples with the same label across data sources. Specifically, it employs a dual-tiered regularization to refine the representation learning, including the Local Consistency Matching and the Complementary Consistency Matching.

For the Local Consistency Matching, it applies consistency in relationships between representations across different clients to guide learning process, which is expressed by

$$\mathcal{L}_A(h_k^{c_1}, h_k^{c_2}, h_k^{c_3}) = \left\|\measuredangle(h_k^{c_1}, h_k^{c_2}, h_k^{c_3}) - \measuredangle(u^{c_1}, u^{c_2}, u^{c_3})\right\|_2, \quad (7)$$

where $h_k^{c_i} = H_g(\mu_k^{c_i} + \triangle)$, if $\mu_k^{c_i}$ is a local prototype of class c_i in the client k, $\triangle = 0$; if $\mu_k^{c_i}$ is an augmented feature, $\triangle = \mathcal{N}(0, \Sigma_{fuse})$. $\measuredangle(h_k^{c_1}, h_k^{c_2}, h_k^{c_3}) = \left\langle \frac{h_k^{c_1} - h_k^{c_2}}{\left\|h_k^{c_1} - h_k^{c_2}\right\|_2}, \frac{h_k^{c_3} - h_k^{c_2}}{\left\|h_k^{c_3} - h_k^{c_2}\right\|_2} \right\rangle$. $\langle\cdot\rangle$ denotes dot product.

Meanwhile, to achieve rapid alignment within limited training epochs, distance-based consistency constraints are also applied, i.e.,

$$\mathcal{L}_E(h_k^{c_1}, h_k^{c_2}) = \left\|dist(h_k^{c_1}, h_k^{c_2}) - dist(u^{c_1}, u^{c_2}))\right\|_2, \quad (8)$$

where $dist(\cdot)$ is an Euclidean distance. Overall, the local matching loss is defined by

$$\mathcal{L}_{LCM} = \sum_{c_i \in \cup_{c=1}^{K}, k \in \cup_{k=1}^{K}} (\mathcal{L}_A(h_k^{c_1}, h_k^{c_1}, h_k^{c_1}) + \mathcal{L}_E(h_k^{c_1}, h_k^{c_2})). \quad (9)$$

For Complementary Consistency Matching, it uses complementary features from diverse sources to help the model learn consistent attributes across clients, which enables the model to transcend the limitations of a single perspective,

$$\mathcal{L}_{CCM} = \sum_{c_i \in \cup_{c=1}^{K}, k_i \in \cup_{k=1}^{K}}^{k_1 \neq k_2 \neq k_3} (\mathcal{L}_A(h_{k_1}^{c_1}, h_{k_2}^{c_2}, h_{k_3}^{c_3}) + \mathcal{L}_E(h_{k_1}^{c_1}, h_{k_2}^{c_2})). \quad (10)$$

Moreover, to enhance robustness and maintain decision boundaries, the CSFSA module uses importance scores s_i learned from clients to downweight lower-quality features, leading to a weighted supervised classification loss, i.e.,

$$\mathcal{L}_{WCE} = -\sum_{i=1}^{N} s_i \left(y_i \log(p_i) + (1 - y_i) \log(1 - p_i)\right). \quad (11)$$

Training Strategies

FedFSA obtains the final model through the training of local client models and the calibration of the global model. It has following training strategies. First, FedFSA aims to calibrate local distributions on the client side, its optimization objective loss is defined as

$$\mathcal{L}_{Client} = \mathcal{L}_{EM} + \alpha\mathcal{L}_{DC}. \quad (12)$$

Furthermore, FedFSA further minimizes distribution discrepancies across different spaces on the server side, the loss for optimization is characterized as

$$\mathcal{L}_{Server} = \mathcal{L}_{WCE} + \eta(\mathcal{L}_{LCM} + \mathcal{L}_{CCM}), \quad (13)$$

where α and η are weighted parameters.

Experiments

Experiment Settings

Datasets. Following existing studies (Li, He, and Song 2021; Luo et al. 2021; Mu et al. 2023), experiments were conducted on three datasets, including CIFAR10 (Krizhevsky, Hinton et al. 2009), CIFAR100 (Krizhevsky, Hinton et al. 2009) and TinyImageNet (Le and Yang 2015) to validate the effectiveness of the FedFSA. The statistical details are presented in the Table 1. And the dataset is partitioned using the Dirichlet distribution with $\beta = 0.5$.

Table 1: Statistics of CIFAR10, CIFAR100 and TinyImagenet datasets used in the experiment.

Datasets	#Class	#Training	#Testing	#Image Size
CIFAR10	10	50000	10000	32 * 32
CIFAR100	100	50000	10000	32 * 32
TinyImagenet	200	100000	10000	64 * 64

Table 2: Performance comparison between FedFSA with existing methods on CIFAR10, CIFAR100 and TinyImagenet datasets.

Methods		CIFAR10		CIFAR100		TinyImagenet	
		K=5	K=10	K=5	K=10	K=5	K=10
Base	FedAvg (AISTATS'17)	70.85	68.24	60.67	57.58	49.58	46.12
Methods based on Knowledge Distillation	MOON (CVPR'21)	71.43	69.44	61.54	58.82	50.12	47.38
	FedProc (FGCS'23)	72.64	69.85	62.04	59.32	50.23	47.79
	FedDeccor (ICLR'23)	72.11	70.21	61.59	59.24	49.75	47.63
	FedETF (ICCV'23)	73.03	70.79	62.36	60.45	50.46	48.25
	FedRCL (CVPR'24)	71.54	69.25	61.48	58.67	50.45	47.46
Methods based on Model Calibration	CCVR (NeurIPS'21)	71.25	69.67	60.67	58.59	49.67	46.23
	FedCSPC (MM'23)	73.24	70.85	62.87	60.88	50.31	48.12
	CLIP2FL (AAAI'24)	72.89	70.49	63.27	61.05	50.74	48.26
	FedFSA$_{FedAvg}$ (Ours)	74.45	72.35	**64.48**	62.41	51.05	48.73
	FedFSA$_{FedETF}$ (Ours)	**75.15**	**72.53**	64.23	**62.58**	**51.42**	**49.16**

Evaluation Measures. Following existing studies (Li, He, and Song 2021; Mu et al. 2023), this study employs the Top-1 Accuracy to evaluate the performance of methods, i.e.,

$$\text{Accuracy } = N_{correct}/N_{total} \tag{14}$$

where $N_{correct}$, N_{total} are the number of correct predictions and total samples, respectively.

Network Architecture. Following existing studies (Li, He, and Song 2021; Mu et al. 2023), the network setup includes an image encoder, a projection head with a 2-layer MLP, and a classifier with a single-layer fully-connected network. We use a CNN with two 5x5 convolutional layers, 2x2 max pooling, and two ReLU-activated fully-connected layers as the encoder on CIFAR10 and use a ResNet18 encoder on other datasets, omitting its last fully-connected layer.

Implementation Details. Following existing studies (Li, He, and Song 2021; Mu et al. 2023), we set clients size $K = 5$ and $K = 10$ in cross-silo settings, the local training epochs $E = 10$, the batch size $B = 64$, the communication round $T = 100$ for CIFAR10 and CIFAR100 datasets, $T = 50$ for TinyImagenet dataset, the learning rate $lr = 0.01$ and the weight decay $wd = 1e - 05$ in the SGD optimizer. The weighted parameter $\alpha = \{0.1, 0.5, 1, 5\}$, the temperature $\tau = 0.5$, the number of clusters $N_v = \{1, 2, 3\}$. The weighted parameter $\eta = \{0.01, 0.1, 1\}$, $\kappa = \{0.3, 0.5, 0.7\}$, the number of augmented features $J = \{1, 2, 4, 8\}$. For other methods, we tuned their hyper-parameters by referring to corresponding papers for fair comparison.

Performance Comparison

We compare FedFSA with nine state-of-the-art methods, including FedAvg (McMahan et al. 2017), MOON (Li, He, and Song 2021), CCVR (Luo et al. 2021), FedProc (Mu et al. 2023), FedDecorr (Shi et al. 2023), FedETF (Li et al. 2023), FedRCL (Seo et al. 2024), FedCSPC (Qi et al. 2023) and CLIP2FL (Shi et al. 2024). The following results can be observed from Table 2.

- **FedFSA is a general framework that can combine various knowledge distillation based approaches**, such as FedAvg and FedETF, to bring them performance gains, which showcases its model-agnostic capability.
- **Model calibration-based methods typically outperform knowledge distillation-based methods, as demonstrated by FedCSPC, CLIP2FL and FedFSA.** This is because they all endeavor to utilize information from multiple sources to train a generalized model.
- **FedETF employs a unified simplex equiangular tight frame classifier often results in better outcomes than methods based on data-driven knowledge (FedProc, FedNTD, MOON)**. This may be due to they avoid issues of poor knowledge quality caused by data disparities and inherent limitations of the models themselves.
- **With an increase in the number of data sources, there is often a decline in the performance**. This results from the amplified disparities across data distributions. FedFSA retains its superiority in performance, fully demonstrating the efficacy of its calibration mechanism.

Ablation Study

This section explores the effectiveness of FedFSA's components with $K = 5$ and $K = 10$ clients, and a Dirichlet parameter $\beta = 0.5$. The results are shown in Table 3.

- **The Text-Enhanced Representation Learning (TERL) module plays a crucial role, contributing an average performance gain of 1.2% to the baseline method**, which verifies that providing unified guidance to different clients aids in enhancing their collaborative outcomes.
- The collaboration between the Cross-Silo Feature Space Alignment (CSFSA) and TERL modules has resulted in a significant improvement in accuracy. **This enhancement has provided an approximate 3% increase for the baseline methods across all cases.**
- **The CSFSA module alone can also produce good results**, as it mitigates the impact of outliers in a weighted manner compared to existing methods.

In-depth Analysis

Robustness of FedFSA on Hyperparameters. This section evaluates the robustness of FedFSA in different hyperparameters. We select the N_v, weight parameters α, η and κ from $\{1, 2, 3\}$, $\{0.1, 0.5, 1, 5\}$, $\{0.01, 0.1, 1\}$ and

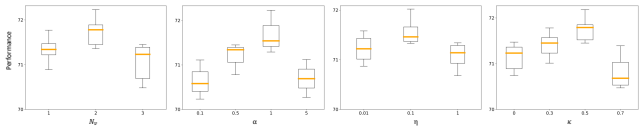


Figure 3: The impact of hyperparameters on performance.

Table 3: Ablation study on the effectiveness of main modules of FedFSA on the CIFAR10 and CIFAR100 datasets.

	CIFAR10		CIFAR100	
	K=5	K=10	K=5	K=10
Base	70.85	68.24	60.67	57.58
+TERL	72.06	70.01	62.53	59.46
+TERL+CSFSA	73.21	71.02	63.26	61.03
+VTSC+CSFSA	73.63	71.21	63.55	61.14
+TERL+VTSC+CSFSA	**74.45**	**72.35**	**64.48**	**62.41**

$\{0, 0.3, 0.5, 0.7\}$. As shown in Figure 3, **FedFSA consistently outperforms FedAvg across various scenarios and demonstrates insensitivity to hyperparameter variations over a wide range**, indicating its strong robustness in hyperparameter selection. Additionally, the model performs best with 2 clusters, as a single cluster may miss intra-class variability, while too many clusters could dilute key features and focus on noise. For α and η, the model performs best when $\alpha = 2$ and $\eta = 0.1$. This is because lower values of α or η might result in the model assigning too little weight to key features, while higher values could lead to over-reliance on certain specific features, ignoring other valuable information. Notably, **fusing an appropriate level of variance knowledge is beneficial, but excessive fusion may lead to inter-class feature overlap, introducing noise and resulting in degraded performance.**

The Effect of the Number of Augmented Features on Performance. This section discusses the effect of augmented feature numbers on calibration results. We adjust N_{aug} from $\{0, 1, 2, 4, 8\}$ with $N_v = 2$. $N_{aug} = 0$ means training with only local prototypes. Figure 4 shows the results. **Increasing the number of augmented features generally improves performance by enriching the feature space and simulating real distributions**, which helps prevent overfitting. Even a few augmented samples can boost performance by about 3%. However, performance on CIFAR100 declines when $N_v = 8$ due to fewer samples per class and overlapping distributions, highlighting the importance of effective feature learning in complex tasks.

Case Study

The Impact of Text-Enhanced Representation Learning on Feature Learning. This section evaluates the impact of Text-Enhanced Representation Learning on feature learning, prototype modeling, and test performance. We selected two clients with different data distributions and used t-SNE to visualize feature distribution for two classes in both training and testing sets. As shown in Figure 5, FedProc and FedFSA learned more discriminative representations compared to FedAvg, especially for majority classes (e.g., the sandy brown class). However, FedProc struggled with classes with few samples due to error accumulation across training rounds. FedFSA uses consistent features to guide local training, ensuring similar representations for shared classes, which helps it outperform other methods. Additionally, FedFSA evaluates prototype significance, assigning low weights to prototypes in overlapping regions of different classes, aiding the CSFSA module in reducing outlier interference.

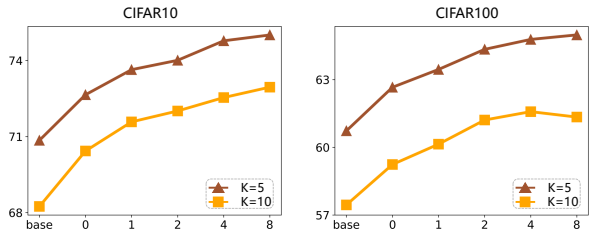


Figure 4: The effect of the number of augmented features $N_f = \{0, 1, 2, 4, 8\}$ on performance of FedFSA on CIFAR10 with different number of clients $K = \{5, 10\}$.

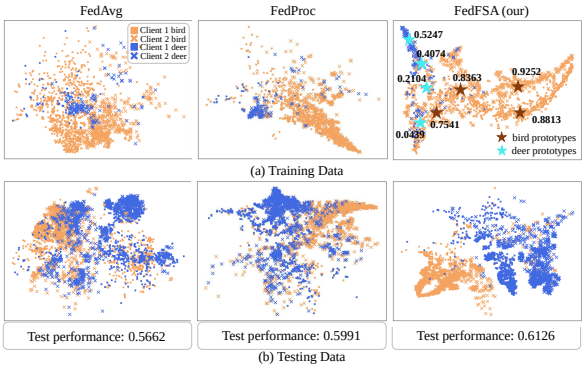
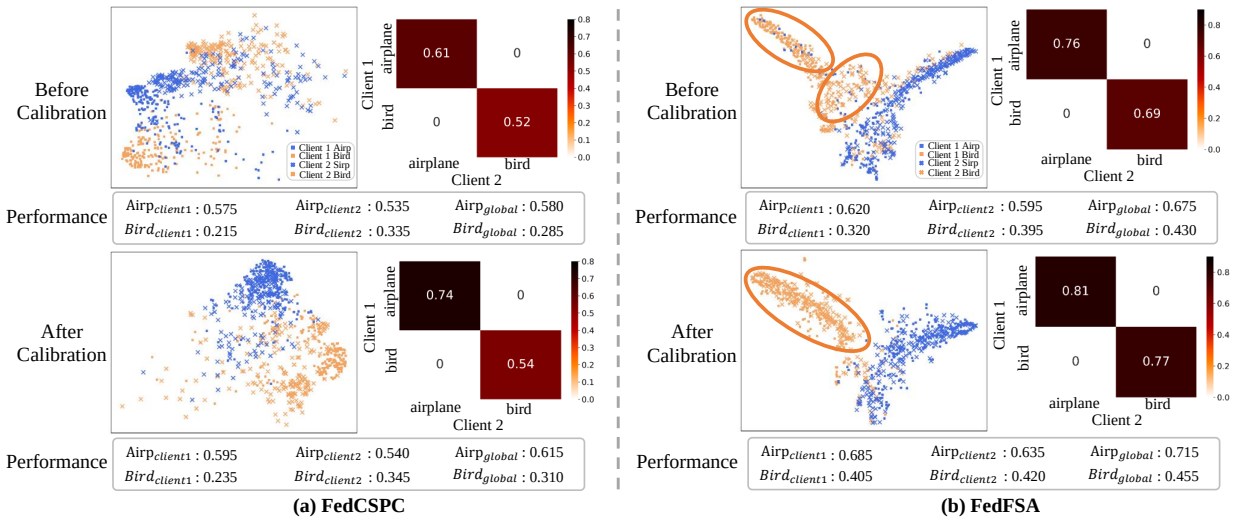


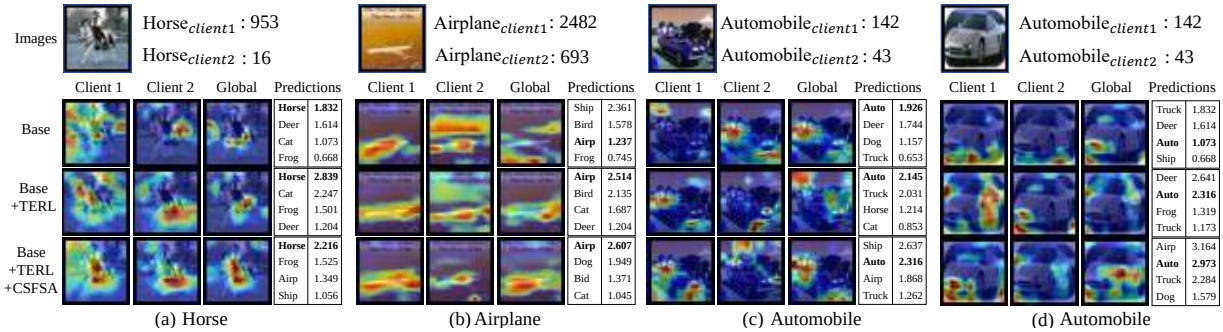
Figure 5: Local feature distributions learned by FedAvg, FedProc and FedFSA on CIFAR10 training and testing set.

Visualization Analysis of Cross-Silo Feature Space Alignment. In this section, we randomly selected two clients and two shared categories (birds with fewer samples and airplanes with more). Figure 6 shows the representation distributions, the CKA similarity (Kornblith et al. 2019; Gao et al. 2024; Liu et al. 2022), and model performance learned from FedCSPC and FedFSA methods. Results indicate that FedFSA learns more compact and discriminative representations within and between classes than FedCSPC. Additionally, FedFSA reduces feature space heterogeneity across clients even before calibration, aiding cross-source feature alignment. This improvement is also reflected in CKA similarity. Conversely, FedCSPC aligns the airplane class better than the bird class due to limited representation quality from minority samples. This is due to poor representation quality from minority samples hindering feature alignment. Furthermore, feature heterogeneity may decrease collaborative performance (see Figure 6(a)). And model calibration typically enhances the local models' personalized capabil-



(a) FedCSPC **(b) FedFSA**

Figure 6: Comparison of representation learning across clients between FedCSPC and FedFSA. FedFSA improves the effectiveness of cross-silo feature alignment for classes with minority samples, and enhances the performance of the global model.



(a) Horse (b) Airplane (c) Automobile (d) Automobile

Figure 7: Error analysis. (a) FedFSA improves the local feature learning and the generalization of global model. (b) FedFSA can calibrate attention towards minority class sample, improving the performance of aggregation. (c) FedFSA failed due to unreliable local learning. (d) FedFSA narrows the gap between actual outcomes and top-1 predictions.

ity, which leverages other clients' knowledge to compensate for shortcomings.

Error analysis. This section uses GradCAM (Selvaraju, Cogswell, and et al 2017; Meng et al. 2019) visualizations to examine FedFSA. Figure 7(a) shows that incorporating TERL and CSFSA modules can improve the base method by using a large number of samples for better feature learning. Figure 7(b) indicates that limited samples may cause base model failure, reducing collaboration effectiveness. Notably, the TERL module can enhance the feature learning, which helps the CSFSA module correct prediction errors. However, model calibration may fail, as shown in Figure 7(c). Despite accurate predictions by the client models, they fail to reliably focus on target regions, resulting in poor representation learning that hinder effective calibration. Finally, Figure 7(d) shows these methods often mispredict classes with few samples. The CSFSA module helps the model focus better, reducing prediction errors. These findings highlight the negative impact of imbalanced data on federated learning and confirm the proposed framework's effectiveness.

Conclusion

To address the issue of ill-posed aggregation caused by data imbalance, this paper proposes a method (FedFSA) for aligning feature spaces across silos. FedFSA introduces the variance transfer technique to construct the prototypical space, which calibrates feature distribution of minority classes. Moreover, FedFSA aligns feature spaces from different sources to bridge inconsistency, fitting data from all clients to obtain a generalized model. Experimental results show that aligning the feature spaces of different clients can improve the performance of the retrained model.

Despite FedFSA mitigates the impact of data imbalance, there are still some directions worth exploring. Firstly, stronger strategies for representation alignment and causal discovery to enhance collaborative modeling (Chen et al. 2023a,b; Wang et al. 2022b,a; Lin et al. 2020; Yang et al. 2024a). Secondly, it makes sense to extend this to more challenging tasks, such as video classification (Wang et al. 2023b, 2024b) and recommendation systems (Ma et al. 2023; Meng et al. 2020).

Acknowledgments

This work is supported in part by the Shandong Province Excellent Young Scientists Fund Program (Overseas) (Grant no. 2022HWYQ-048), the TaiShan Scholars Program (Grant no. tsqn202211289), the National Natural Science Foundation of China (NSFC) Joint Fund Key Project (Grant No. U2336211).

References

Cai, J.; Zhang, Y.; Fan, J.; and Ng, S.-K. 2024a. LG-FGAD: An Effective Federated Graph Anomaly Detection Framework. In *IJCAI*.

Cai, J.; Zhang, Y.; Lu, Z.; Guo, W.; and Ng, S.-k. 2024b. Towards Effective Federated Graph Anomaly Detection via Self-boosted Knowledge Distillation. In *MM*, 5537–5546.

Chen, Y.; Tan, A. Z.; Feng, S.; Yu, H.; Deng, T.; Zhao, L.; and Wu, F. 2024. General Federated Class-Incremental Learning With Lightweight Generative Replay. *IEEE Internet of Things Journal*, 11: 33927–33939.

Chen, Z.; Qi, Z.; Cao, X.; Li, X.; Meng, X.; and Meng, L. 2023a. Class-level Structural Relation Modeling and Smoothing for Visual Representation Learning. In *MM*, 2964–2972.

Chen, Z.; Qi, Z.; Li, X.; Wang, Y.; Meng, L.; and Meng, X. 2023b. Class-aware convolution and attentive aggregation for image classification. In *MM Asia*, 1–7.

Gao, Y.; Hou, Z.; Yang, C.; Li, Z.; Yu, H.; and Li, X. 2024. The Prospect of Enhancing Large-Scale Heterogeneous Federated Learning with Foundation Models. In *ICME*, 1–6.

Hu, M.; Cao, Y.; Li, A.; Li, Z.; Liu, C.; Li, T.; Chen, M.; and Liu, Y. 2024a. FedMut: Generalized Federated Learning via Stochastic Mutation. In *AAAI*, volume 38, 12528–12537.

Hu, M.; Yue, Z.; Xie, X.; Chen, C.; Huang, Y.; Wei, X.; Lian, X.; Liu, Y.; and Chen, M. 2024b. Is Aggregation the Only Choice? Federated Learning via Layer-wise Model Recombination. In *KDD*, 1096–1107.

Hu, M.; Zhou, P.; Yue, Z.; Ling, Z.; Huang, Y.; Li, A.; Liu, Y.; Lian, X.; and Chen, M. 2024c. FedCross: Towards Accurate Federated Learning via Multi-Model Cross-Aggregation. In *ICDE*, 2137–2150.

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.

Kairouz, P.; McMahan, H. B.; Avent, B.; Bellet, A.; Bennis, M.; Bhagoji, A. N.; Bonawitz, K.; Charles, Z.; Cormode, G.; Cummings, R.; et al. 2021. Advances and open problems in federated learning. *Foundations and trends® in machine learning*, 14(1–2): 1–210.

Kornblith, S.; Norouzi, M.; Lee, H.; and Hinton, G. 2019. Similarity of neural network representations revisited. In *ICML*, 3519–3529.

Krizhevsky, A.; Hinton, G.; et al. 2009. Learning multiple layers of features from tiny images.

Le, Y.; and Yang, X. 2015. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7): 3.

Lee, G.; Jeong, M.; Shin, Y.; Bae, S.; and Yun, S.-Y. 2022. Preservation of the global knowledge by not-true distillation in federated learning. In *NeurIPS*, volume 35, 38461–38474.

Li, Q.; He, B.; and Song, D. 2021. Model-contrastive federated learning. In *CVPR*, 10713–10722.

Li, X.; Ma, H.; Meng, L.; and Meng, X. 2021. Comparative study of adversarial training methods for long-tailed classification. In *ADVM*, 1–7.

Li, X.; Zheng, Y.; Ma, H.; Qi, Z.; Meng, X.; and Meng, L. 2024a. Cross-modal learning using privileged information for long-tailed image classification. *Computational Visual Media*, 1–12.

Li, Z.; Shang, X.; He, R.; Lin, T.; and Wu, C. 2023. No fear of classifier biases: Neural collapse inspired federated learning with synthetic and fixed classifier. In *ICCV*, 5319–5329.

Li, Z.; Sun, Y.; Shao, J.; Mao, Y.; Wang, J. H.; and Zhang, J. 2024b. Feature matching data synthesis for non-iid federated learning. *IEEE Transactions on Mobile Computing*, 23(10): 9352–9367.

Lin, C.; Zhao, S.; Meng, L.; and Chua, T.-S. 2020. Multi-source domain adaptation for visual sentiment classification. In *AAAI*, volume 34, 2661–2668.

Lindsay, B. G. 1995. Mixture models: theory, geometry, and applications. Ims.

Liu, T.; Qi, Z.; Chen, Z.; Meng, X.; and Meng, L. 2023. Cross-Training with Prototypical Distillation for improving the generalization of Federated Learning. In *ICME*, 648–653.

Liu, Z.; Chen, Y.; Yu, H.; Liu, Y.; and Cui, L. 2022. Gtg-shapley: Efficient and accurate participant contribution evaluation in federated learning. *ACM Transactions on intelligent Systems and Technology*, 13(4): 1–21.

Luo, M.; Chen, F.; Hu, D.; Zhang, Y.; Liang, J.; and Feng, J. 2021. No fear of heterogeneity: Classifier calibration for federated learning with non-iid data. In *NeurIPS*, volume 34, 5972–5984.

Ma, H.; Qi, Z.; Dong, X.; Li, X.; Zheng, Y.; Meng, X.; and Meng, L. 2023. Cross-modal content inference and feature enrichment for cold-start recommendation. In *IJCNN*, 1–8.

McMahan, B.; Moore, E.; Ramage, D.; and et al. 2017. Communication-efficient learning of deep networks from decentralized data. In *AISTATS*, 1273–1282.

Meng, L.; Chen, L.; Yang, X.; Tao, D.; Zhang, H.; Miao, C.; and Chua, T.-S. 2019. Learning using privileged information for food recognition. In *MM*, 557–565.

Meng, L.; Feng, F.; He, X.; Gao, X.; and Chua, T.-S. 2020. Heterogeneous fusion of semantic and collaborative information for visually-aware food recommendation. In *MM*, 3460–3468.

Meng, L.; Qi, Z.; Wu, L.; Du, X.; Li, Z.; Cui, L.; and Meng, X. 2024. Improving Global Generalization and Local Personalization for Federated Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 1–12.

Meng, L.; Tan, A.-H.; and Miao, C. 2019. Salience-aware adaptive resonance theory for large-scale sparse data clustering. *Neural Networks*, 120: 143–157.

Meng, L.; Tan, A.-H.; and Wunsch, D. C. 2015. Adaptive scaling of cluster boundaries for large-scale social media data clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 27(12): 2656–2669.

Mu, X.; Shen, Y.; Cheng, K.; Geng, X.; Fu, J.; Zhang, T.; and Zhang, Z. 2023. Fedproc: Prototypical contrastive federated learning on non-iid data. *Future Generation Computer Systems*, 143: 93–104.

Nie, L.; Zhang, L.; Meng, L.; Song, X.; Chang, X.; and Li, X. 2016. Modeling disease progression via multisource multitask learners: A case study with Alzheimer's disease. *IEEE Transactions on Neural Networks and Learning Systems*, 28(7): 1508–1519.

Qi, Z.; He, W.; Meng, X.; and Meng, L. 2024a. Attentive modeling and distillation for out-of-distribution generalization of federated learning. In *ICME*, 1–6.

Qi, Z.; Meng, L.; Chen, Z.; Hu, H.; Lin, H.; and Meng, X. 2023. Cross-Silo Prototypical Calibration for Federated Learning with Non-IID Data. In *MM*, 3099–3107.

Qi, Z.; Meng, L.; He, W.; Zhang, R.; Wang, Y.; Qi, X.; and Meng, X. 2024b. Cross-Training with Multi-View Knowledge Fusion for Heterogenous Federated Learning. *arXiv preprint arXiv:2405.20046*.

Qi, Z.; Wang, Y.; Chen, Z.; Wang, R.; Meng, X.; and Meng, L. 2022. Clustering-based curriculum construction for sample-balanced federated learning. In *CICAI*, 155–166.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. 2021. Learning transferable visual models from natural language supervision. In *ICML*, 8748–8763.

Ren, C.; Yu, H.; Peng, H.; Tang, X.; Li, A.; Gao, Y.; Tan, A. Z.; Zhao, B.; Li, X.; Li, Z.; et al. 2024. Advances and open challenges in federated learning with foundation models. *arXiv preprint arXiv:2404.15381*.

Selvaraju, R. R.; Cogswell, M.; and et al. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *ICCV*, 618–626.

Seo, S.; Kim, J.; Kim, G.; and Han, B. 2024. Relaxed contrastive learning for federated learning. In *CVPR*, 12279–12288.

Shang, X.; Lu, Y.; Huang, G.; and Wang, H. 2022. Federated learning on heterogeneous and long-tailed data via classifier re-training with federated features. *preprint arXiv:2204.13399*.

Shi, J.; Zheng, S.; Yin, X.; Lu, Y.; Xie, Y.; and Qu, Y. 2024. CLIP-Guided Federated Learning on Heterogeneity and Long-Tailed Data. In *AAAI*, volume 38, 14955–14963.

Shi, Y.; Liang, J.; Zhang, W.; Tan, V. Y.; and Bai, S. 2023. Towards Understanding and Mitigating Dimensional Collapse in Heterogeneous Federated Learning. In *ICLR*.

Tan, Y.; Long, G.; Liu, L.; Zhou, T.; Lu, Q.; Jiang, J.; and Zhang, C. 2022. Fedproto: Federated prototype learning across heterogeneous clients. In *AAAI*, volume 36, 8432–8440.

Wang, H.; Li, Y.; Xu, W.; Li, R.; Zhan, Y.; and Zeng, Z. 2023a. Dafkd: Domain-aware federated knowledge distillation. In *CVPR*, 20412–20421.

Wang, H.; Zheng, P.; Han, X.; Xu, W.; Li, R.; and Zhang, T. 2024a. FedNLR: Federated Learning with Neuron-wise Learning Rates. In *KDD*, 3069–3080.

Wang, L.; Xu, S.; Wang, X.; and Zhu, Q. 2021. Addressing class imbalance in federated learning. In *AAAI*, volume 35, 10165–10173.

Wang, Y.; Li, X.; Ma, H.; Qi, Z.; Meng, X.; and Meng, L. 2022a. Causal inference with sample balancing for out-of-distribution detection in visual classification. In *CICAI*, 572–583.

Wang, Y.; Li, X.; Qi, Z.; Li, J.; Li, X.; Meng, X.; and Meng, L. 2022b. Meta-causal feature learning for out-of-distribution generalization. In *ECCV*, 530–545.

Wang, Y.; Meng, L.; Ma, H.; Wang, Y.; Huang, H.; and Meng, X. 2024b. Modeling Event-level Causal Representation for Video Classification. In *MM*, 3936–3944.

Wang, Y.; Qi, Z.; Li, X.; Liu, J.; Meng, X.; and Meng, L. 2023b. Multi-channel attentive weighting of visual frames for multimodal video classification. In *IJCNN*, 1–8.

Wen, J.; Zhang, Z.; Lan, Y.; Cui, Z.; Cai, J.; and Zhang, W. 2023. A survey on federated learning: challenges and applications. *International Journal of Machine Learning and Cybernetics*, 14(2): 513–535.

Wu, N.; Yu, L.; Yang, X.; Cheng, K.-T.; and Yan, Z. 2023. FedIIC: Towards robust federated learning for class-imbalanced medical image classification. In *MICCAI*, 692–702.

Yang, X.; Chang, T.; Zhang, T.; Wang, S.; Hong, R.; and Wang, M. 2024a. Learning Hierarchical Visual Transformation for Domain Generalizable Visual Matching and Recognition. *International Journal of Computer Vision*, 1–27.

Yang, Z.; Zhang, Y.; Zheng, Y.; Tian, X.; Peng, H.; Liu, T.; and Han, B. 2024b. FedFed: Feature distillation against data heterogeneity in federated learning. In *NeurIPS*, volume 36, 60397–60428.

Ye, R.; Ni, Z.; Xu, C.; Wang, J.; Chen, S.; and Eldar, Y. C. 2023. Fedfm: Anchor-based feature matching for data heterogeneity in federated learning. *IEEE Transactions on Signal Processing*, 71: 4224–4239.

Yu, F.; Zhang, W.; Qin, Z.; Xu, Z.; Wang, D.; Liu, C.; Tian, Z.; and Chen, X. 2021. Fed2: Feature-aligned federated learning. In *KDD*, 2066–2074.

Zeng, Y.; Liu, L.; Liu, L.; Shen, L.; Liu, S.; and Wu, B. 2023. Global Balanced Experts for Federated Long-Tailed Learning. In *ICCV*, 4815–4825.

Zhang, L.; Shen, L.; Ding, L.; Tao, D.; and Duan, L.-Y. 2022. Fine-tuning global model via data-free knowledge distillation for non-iid federated learning. In *CVPR*, 10174–10183.