FEDSP: FEDERATED SHARED-PERSONALIZED LEARNING FOR DOMAIN GENERALIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Federated Domain Generalization (FedDG) aims to learn models that generalize to unseen domains from distributed client data without accessing clients' raw data. However, existing federated learning methods typically promote collaboration by aggregating client models into a shared global model, implicitly assuming that knowledge beneficial to one client can be effectively shared across all clients. In the presence of domain shift, client-specific visual variations may become entangled with domain-invariant semantic knowledge, making it challenging for the model to determine which knowledge should be shared and which should be personalized. To address this challenge, we propose Federated Shared-Personalized Learning for Domain Generalization (FedSP), which explicitly decouples shared knowledge from client-specific knowledge during federated learning. FedSP employs category prototype-driven cross-domain semantic alignment in the shared module and residual adaptation of domain-inherent discriminative attributes in the personalized module to learn transferable domain-invariant shared representations while modeling domain-specific semantic variations. Additionally, FedSP introduces a multi-scale uncertainty-constrained learning strategy to enhance the model's adaptability across diverse domain shifts. Extensive experiments on multiple domain generalization benchmarks demonstrate that FedSP consistently improves model generalization to unseen domains and outperforms existing federated learning and domain generalization methods. Further analysis shows that explicitly decoupling shared and personalized knowledge facilitates more effective federated collaboration and enhances model robustness under significant domain shifts.

1 INTRODUCTION

As machine vision models are increasingly deployed in real-world applications, they are often exposed to data distributions that differ from those encountered during training, resulting in substantial performance degradation under distribution shifts (Al-Emadi et al., 2025; Li et al., 2025a). This has made robust generalization to unseen conditions an important objective beyond conventional in-distribution performance. Domain generalization (DG) addresses this challenge by learning from multiple source domains and seeking domain-stable visual representations that enable models to maintain reliable classification performance on unseen target domains (Blanchard et al., 2011; Muandet et al., 2013; Zhou et al., 2023; Ballas & Diou, 2025). However, existing DG methods typically assume that data from multiple source domains can be centrally accessed and jointly used for training. In practical applications, data collected across different locations, devices, or scenarios are often distributed across isolated sources and subject to constraints such as data privacy, communication costs, and data ownership, making direct data aggregation infeasible (McMahan et al., 2017; Kairouz et al., 2021). This raises a fundamental challenge for domain generalization in image classification: how to learn domain-stable visual representations when data cannot be directly shared across domains.

Federated Learning (FL) provides an effective paradigm for collaborative model training under decentralized data constraints and has recently been increasingly adopted for domain generalization, giving rise to federated domain generalization (FedDG), which seeks to leverage heterogeneous data distributed across clients to learn models that generalize to unseen domains (McMahan et al., 2017; Kairouz et al., 2021; Li et al., 2025b; Gupta et al., 2025; Wang et al., 2026). Existing FedDG

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

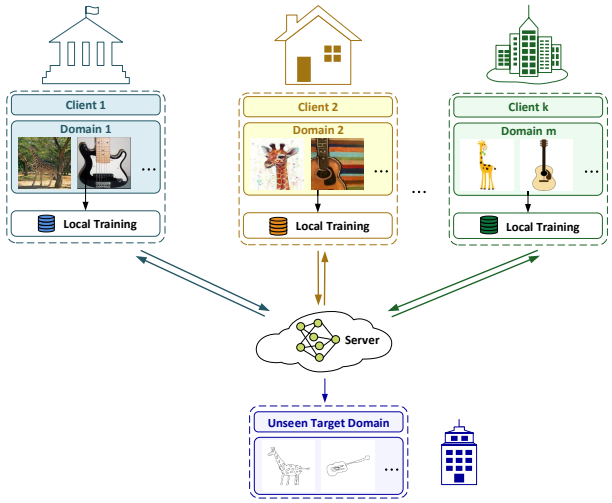


Figure 1: Overview of the Federated Domain Generalization Task.

methods primarily facilitate cross-client knowledge sharing through local model training and global parameter aggregation, with recent studies exploring strategies such as cross-client feature alignment, style diversification, and aggregation optimization to alleviate distribution discrepancies and improve generalization to unseen domains (Zhang et al., 2023b; Chen et al., 2023; 2024; Gupta et al., 2025). However, distributional differences may arise from either domain-specific factors irrelevant to classification or semantic variations critical for discrimination. Treating both uniformly as heterogeneity to eliminate through parameter aggregation or feature alignment may suppress domain biases while discarding discriminative semantic information (Ding et al., 2022), leaving personalized information with discriminative value insufficiently exploited.

To address these issues, we propose FedSP, which tackles federated domain generalization from three perspectives: what to share, what to adapt, and how to constrain training. We introduce category prototype collaboration that aggregates prototypes from different clients and employs contrastive constraints between same-class prototypes to establish cross-domain semantic connections, reducing interference from local domain features (Tan et al., 2022). To accommodate client-specific variations while preserving discriminative structure, we design residual feature adaptation that maintains original representations. Since clients contribute differently to global optimization and simple parameter averaging may fail to capture actual contributions (Li et al., 2021b; Chen & Chao, 2021; Acar et al., 2021), we further introduce an uncertainty-aware mechanism based on evidential learning and predictive entropy to characterize client reliability (Sensoy et al., 2018) and dynamically adjust their contributions. Through these complementary mechanisms, FedSP learns shared semantics, adapts to local variations, and coordinates client contributions for stable cross-domain representations.

The main contributions of this work are summarized as follows:

(1) We propose Federated Shared-Personalized Learning for Domain Generalization (FedSP), a federated domain generalization framework for distributed application scenarios that enables efficient collaborative learning across clients.

(2) We design the Category Prototype-Driven Cross-Domain Semantic Alignment and Sharing Module (CSAS), the Residual Adaptation of Domain-Specific Discriminative Attributes (DSRA) module, and the Systematic Multi-Scale Uncertainty-Constrained Learning (MUCL) method, which respectively address what should be shared, what should be adapted, and how learning should be constrained across clients in a federated learning system.

(3) Extensive experiments on PACS, Office-Home, Office-Caltech10, and Terra Incognita demonstrate that FedSP achieves state-of-the-art performance, with ablation studies validating the effectiveness of each component.

2 Related Work

2.1 Domain Generalization

Recent studies on domain generalization have primarily improved cross-domain generalization through data augmentation (Zhou et al., 2021; Li et al., 2021a; Kim et al., 2023), feature representation (Kim et al., 2021; Bui et al., 2021; Ding et al., 2022), and model optimization (Ramé et al., 2022; Cha et al., 2021; Krueger et al., 2021; Sener & Koltun, 2022; Wang et al., 2023). Data augmentation methods simulate domain variations to increase training diversity, such as MixStyle (Zhou et al., 2021), which mixes feature statistics to generate novel styles. Feature representation methods like SelfReg (Kim et al., 2021) enhance robustness through self-supervised contrastive regularization. Optimization-based methods constrain cross-domain invariance, such as Fishr (Ramé et al., 2022), which matches gradient variances across domains, and SWAD (Cha et al., 2021), which seeks flatter parameter minima. However, these methods assume centralized source-domain data for computing cross-domain statistics or constructing joint batches. When data are distributed across clients, class semantics and feature distributions cannot be directly shared, making these methods difficult to apply in federated settings.

2.2 Federated Domain Generalization

Recent research on federated domain generalization has focused on data distribution expansion (Chen et al., 2023; Park et al., 2023; Yang et al., 2024a;b), global optimization (Zhang et al., 2023b; Nguyen et al., 2025; Zhou et al., 2025; Yan & Guo, 2024; Pourpanah et al., 2025; Wang et al., 2026), and feature representation learning (Le et al., 2024; Li et al., 2024; Wu et al., 2025). Early work enhanced generalization by simulating domain shifts through frequency-domain interpolation and episodic training (Liu et al., 2021) or federated adversarial learning (Zhang et al., 2023a). To improve global optimization, GA optimizes the global objective through variance constraints and dynamic aggregation (Zhang et al., 2023b); FedIIR models implicit invariant relationships from prediction discrepancies and suppresses domain-related updates through gradient alignment (Guo et al., 2023); FedOMG seeks consistent optimization directions at the server to alleviate biases caused by heterogeneous data (Nguyen et al., 2025); and DI dynamically adjusts aggregation weights according to data influence (Zhou et al., 2025). Recent studies have focused more on feature representation and personalized learning. StableFDG learns domain-stable representations through style sharing, style exploration, and attention enhancement (Park et al., 2023); gPerXAN reduces local domain bias through personalized normalization (Le et al., 2024); FedAlign promotes domain-invariant learning through cross-client representation augmentation and two-stage alignment (Gupta et al., 2025); and FedDIM enhances invariance by enforcing consistency between sample-level and class-level decision representations (Liao et al., 2025). However, existing methods still struggle to sufficiently distinguish shared class-level commonalities across domains from domain-specific discriminative information, while differences in client learning reliability remain insufficiently considered. How to achieve complementary learning of shared and personalized knowledge in federated collaboration, while leveraging uncertainty to coordinate client learning and global knowledge aggregation, is the problem addressed in this work.

3 Method

3.1 Problem Definition and Overall Framework

Consider a federated learning system with K clients and a server S, where client k has N_k local samples and the classification task involves C classes. The system accesses M source-domain datasets $\{D_m\}_{m=1}^{M}$. A source dataset may be distributed across multiple clients, while each client accesses only one source dataset. During training, clients optimize their local models using accessible source-domain data and communicate with the server, which performs knowledge aggregation without accessing raw data (McMahan et al., 2017). The objective is to learn a global model that generalizes to an unseen target domain D_u, where D_u is inaccessible during training and $u \neq m$.

The overall framework of FedSP is shown in Figure 2. It consists of client-side learning and server-side collaborative aggregation. To decouple shared knowledge from domain-specific information in

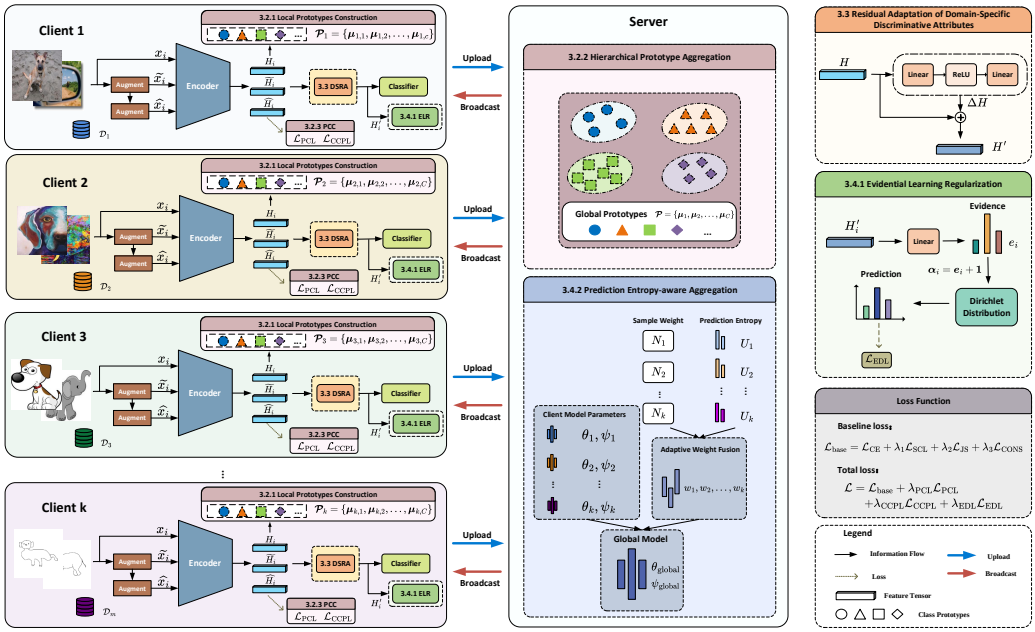


Figure 2: Overall Framework of FedSP.

federated domain generalization, FedSP introduces three complementary modules: CSAS for cross-client category semantic alignment and sharing, DSRA for residual modeling of domain-specific discriminative attributes, and MUCL for uncertainty-constrained learning. Together, these modules enable semantic knowledge sharing, domain-specific information modeling, and reliable collaborative learning to improve generalization to unseen target domains.

3.2 CSAS Module

CSAS consists of local prototype construction, hierarchical prototype aggregation, and prototype contrastive constraints. Clients construct category prototypes and calculate intra-class variance, while the server aggregates global prototypes based on sample counts, variance, and cross-domain consistency. Global and local prototypes guide category structure constraints and cross-domain feature alignment.

3.2.1 Local Prototype Construction (LPC)

Let the local dataset of client k be $\mathcal{D}_k$, $k = 1, \ldots, K$. For each sample x_i, an original view and two augmented views, $\widetilde{x}_i$ and $\widehat{x}_i$, are constructed (Gupta et al., 2025). These views are processed by the shared encoder to obtain the corresponding features H_i, $\widetilde{H}_i$, and $\widehat{H}_i$. Based on these features, the prototype of class c is calculated as:

$$\mu_{k,c} = \frac{1}{2N_k^c} \sum_{i:x_i \in c} \left(H_i + \widetilde{H}_i \right), \tag{1}$$

where N_k^c denotes the number of samples belonging to class c in client k. To measure prototype stability, we compute the intra-class variance across D feature dimensions by averaging the variance of each dimension j:

$$v_{k,c} = \frac{1}{D} \sum_{j=1}^{D} \frac{1}{2N_k^c} \sum_{i:x_i \in c} \left[(H_{i,j} - \mu_{k,c,j})^2 + \left(\widetilde{H_{i,j}} - \mu_{k,c,j} \right)^2 \right]. \tag{2}$$

Each client constructs a local class prototype bank $\mathcal{P}_k = \{\mu_{k,c}\}_{c=1}^{C}$. The client then uploads $\mathcal{P}_k$, $\{N_k^c\}_{c=1}^{C}$, and $\{v_{k,c}\}_{c=1}^{C}$ to the server.

3.2.2 HIERARCHICAL PROTOTYPE AGGREGATION (HPA)

After receiving $\mathcal{P}_k$, $\{N_k^c\}_{c=1}^C$, and $\{v_{k,c}\}_{c=1}^C$ from each client, the server adopts hierarchical aggregation from intra-domain to cross-domain. Let $\mathcal{K}_m$ denote clients in domain $\mathcal{D}_m$. The domain-level class prototype aggregates clients with larger sample counts and smaller variance:

$$\mu_{m,c} = \sum_{k \in \mathcal{K}_{m,c}} \frac{N_k^c / (v_{k,c} + \varepsilon)}{\sum_{j \in \mathcal{K}_{m,c}} N_j^c / (v_{j,c} + \varepsilon)} \mu_{k,c}, \tag{3}$$

where $\mathcal{K}_{m,c}$ denotes the set of clients in domain $\mathcal{D}_m$ that contain class c, and ε is a smoothing term.

Domain-level reliability is defined as the inverse average distance between client and domain prototypes:

$$\rho_{m,c} = \left(\frac{1}{|\mathcal{K}_m|} \sum_{k \in \mathcal{K}_m} ||\mu_{k,c} - \mu_{m,c}||_2^2 + \varepsilon \right)^{-1}. \tag{4}$$

Let $\mathcal{I}_c$ denote source domains containing class c. The inter-domain score $s_{m,c}$ combines domain reliability and cross-domain prototype similarity:

$$s_{m,c} = \frac{\rho_{m,c}}{\sum_{m' \in \mathcal{I}_c} \rho_{m',c}} \cdot \frac{1}{2} \left(1 + \frac{1}{|\mathcal{I}_c| - 1} \sum_{m' \in \mathcal{I}_c, m' \neq m} \cos(\mu_{m,c}, \mu_{m',c}) \right), \tag{5}$$

which is mapped to aggregation weights:

$$w_{m,c} = (1 - \zeta) \frac{\exp(s_{m,c}/\tau_a)}{\sum_{m' \in \mathcal{I}_c} \exp(s_{m',c}/\tau_a)} + \frac{\zeta}{|\mathcal{I}_c|}, \tag{6}$$

where τ_a and ζ are temperature and shrinkage parameters.

The global prototypes are updated via momentum:

$$\mu_c^{(r)} = \eta_p \mu_c^{(r-1)} + (1 - \eta_p) \sum_{m \in \mathcal{I}_c} w_{m,c} \mu_{m,c}, \tag{7}$$

where η_p denotes the prototype momentum coefficient. The updated global class prototype bank $\mathcal{P} = \{\mu_c\}_{c=1}^C$ is distributed with global parameters for next-round training.

3.2.3 PROTOTYPE CONTRASTIVE CONSTRAINTS (PCC)

After receiving the global prototype bank $\mathcal{P} = \{\mu_c\}_{c=1}^C$, clients perform contrastive learning with features H_i, $\widetilde{H}_i$, and $\widehat{H}_i$. Let $\mathcal{C}_p$ denote classes with available prototypes. The prototype contrastive loss is:

$$\ell_{\mathrm{PCL}}(H, y_i) = -\log \frac{\exp(\cos(H, \mu_{y_i})/\tau_p)}{\sum_{c \in \mathcal{C}_p} \exp(\cos(H, \mu_c)/\tau_p)}, \quad H \in \left\{ H_i, \widetilde{H}_i, \widehat{H}_i \right\}, \tag{8}$$

where τ_p is the temperature parameter and y_i is the class label. This loss promotes intra-class compactness and inter-class separation. Multi-view features are aggregated as:

$$\mathcal{L}_{\mathrm{PCL}} = (1 - \lambda) \ell_{\mathrm{PCL}}(H_i, y_i) + \lambda \frac{\ell_{\mathrm{PCL}}\left(\widetilde{H}_i, y_i\right) + \ell_{\mathrm{PCL}}\left(\widehat{H}_i, y_i\right)}{2}. \tag{9}$$

Clients also receive the local class prototype banks uploaded by other clients. Let P_i^+ denote same-class prototypes from other domains and Q_i all cross-domain prototypes. The cross-client prototype contrastive loss is:

$$\ell_{\mathrm{CCPL}}(H, y_i) = -\frac{1}{|P_i^+|} \sum_{\mu \in P_i^+} \log \frac{\exp(\cos(H, \mu)/\tau_c)}{\sum_{\nu \in Q_i} \exp(\cos(H, \nu)/\tau_c)}, \quad H \in \left\{ \widetilde{H}_i, \widehat{H}_i \right\}, \tag{10}$$

where τ_c is the temperature parameter. This loss aligns augmented features with same-class prototypes from different domains while maintaining inter-class boundaries. The final loss is:

$$\mathcal{L}_{\mathrm{CCPL}} = \frac{\ell_{\mathrm{CCPL}}\left(\widetilde{H}_i, y_i\right) + \ell_{\mathrm{CCPL}}\left(\widehat{H}_i, y_i\right)}{2}. \tag{11}$$

3.3 DSRA MODULE

Although category prototype-driven alignment extracts shared class semantics across clients, samples from different domains possess distinctive discriminative attributes that provide additional information for recognition. DSRA captures these client-specific features through residual learning to complement shared semantics.

For features H_i, $\widetilde{H_i}$, and $\widehat{H_i}$ from the backbone, DSRA employs a shared MLP to learn domain-specific attributes and injects them as residuals:

$$\Delta H = W_2 \cdot \mathrm{ReLU}\left(W_1 H + b_1\right) + b_2, \qquad H' = H + \alpha \cdot \Delta H, \qquad H \in \left\{H_i, \widetilde{H}_i, \widehat{H}_i\right\}, \quad (12)$$

where H' is the adapted feature, W_1, W_2, b_1, b_2 are MLP parameters, and α is a learnable residual strength constrained to a small range.After local training, clients upload DSRA parameters $\psi_k^{(r)}$ to the server, which aggregates them via momentum:

$$\psi^{(r)} = \eta_\psi \psi^{(r-1)} + (1 - \eta_\psi) \sum_{k=1}^{K} w_k^{(r)} \psi_k^{(r)}, \quad (13)$$

where $w_k^{(r)}$ is the aggregation weight defined in Section 3.4.2 and η_ψ is the momentum coefficient.

3.4 MUCL MODULE

To address aggregation weight imbalance caused by varying client fitting quality, MUCL establishes uncertainty-aware training and aggregation through client-side evidential regularization and server-side predictive entropy-aware aggregation.

3.4.1 EVIDENTIAL LEARNING REGULARIZATION (ELR)

A lightweight Evidential Learning Head is introduced as an auxiliary regularization branch to convert the output H_i' of DSRA into non-negative class evidence, based on which a Dirichlet distribution is constructed to characterize the degree of support for each class. During training, the standard evidential loss (Sensoy et al., 2018), consisting of a Bayesian risk term and a KL regularization term toward a uniform Dirichlet distribution, is adopted:

$$\mathcal{L}_{\mathrm{EDL}} = \mathcal{L}_{\mathrm{risk}} + \lambda_{\mathrm{KL}} \eta_r \mathcal{L}_{\mathrm{KL}}, \quad (14)$$

where $\mathcal{L}_{\mathrm{risk}}$ constrains the prediction risk of the correct class, $\mathcal{L}_{\mathrm{KL}}$ suppresses an overly concentrated evidence distribution, λ_{KL} denotes the weight of the KL regularization term, and η_r is an annealing coefficient that gradually increases with the training round r to reduce the regularization strength during the early stages of training.

The head participates only in local training without server aggregation, introducing no communication overhead while constraining backbone features toward conservative confidence distributions through backpropagation.

3.4.2 PREDICTIVE ENTROPY-AWARE AGGREGATION (PEA)

After client-side evidential regularization, the server leverages predictive entropy to characterize the discrepancy between the global model and clients. Before round r training, the global model parameters $\{\theta_{\mathrm{global}}^{(r)}, \psi_{\mathrm{global}}^{(r)}\}$ are saved and used post-training to predict on local data:

$$p_{i,c}^{(r)} = \left[\mathrm{softmax}\left(F\left(x_i; \theta_{\mathrm{global}}^{(r)}, \psi_{\mathrm{global}}^{(r)}\right)\right)\right]_c, \quad (15)$$

where $F\left(\cdot; \theta_{\mathrm{global}}^{(r)}, \psi_{\mathrm{global}}^{(r)}\right)$ denotes the complete forward mapping instantiated by the global model parameters. Client k's average predictive entropy is:

$$U_k^{(r)} = -\frac{1}{N_k} \sum_{i=1}^{N_k} \sum_{c=1}^{C} p_{i,c}^{(r)} \ln\left(p_{i,c}^{(r)} + \varepsilon\right), \quad (16)$$

where $U_k^{(r)}$ characterizes the average predictive uncertainty of the current global model on the data of client k.

Based on this, the client predictive entropy is converted into an uncertainty-aware weight and fused with the weight based on the number of local samples in FedAvg (McMahan et al., 2017) to obtain the final aggregation weight of client k:

$$w_k^{(r)} = (1 - \lambda_e) \frac{N_k}{\sum_{j=1}^{K} N_j} + \lambda_e \frac{\exp\left(U_k^{(r)}/\tau_e\right)}{\sum_{j=1}^{K} \exp\left(U_j^{(r)}/\tau_e\right)}, \tag{17}$$

where τ_e is the predictive entropy temperature parameter, and λ_e is the fusion coefficient. Clients with higher uncertainty receive larger weights, guiding global model updates toward better adaptation. The global model is then aggregated as:

$$\theta_{\text{global}}^{(r)} = \sum_{k=1}^{K} w_k^{(r)} \cdot \theta_k^{(r)}. \tag{18}$$

DSRA parameters $\psi_{\text{global}}^{(r)}$ follow Section 3.3 with the same weights and momentum strategy.

3.5 Training Objective

The base training objective $\mathcal{L}_{\text{base}}$ consists of the cross-entropy loss, supervised contrastive loss, prediction consistency loss, and feature consistency loss (Gupta et al., 2025), and is defined as:

$$\mathcal{L}_{\text{base}} = \mathcal{L}_{\text{CE}} + \lambda_1 \mathcal{L}_{\text{SCL}} + \lambda_2 \mathcal{L}_{\text{JS}} + \lambda_3 \mathcal{L}_{\text{CONS}}, \tag{19}$$

where λ_1, λ_2, and λ_3 are loss weights.

Based on this, the training objective of client k in communication round r is defined as:

$$\mathcal{L}_k^{(r)} = \mathcal{L}_{\text{base}} + \lambda_{\text{PCL}}^{(r)} \mathcal{L}_{\text{PCL}} + \lambda_{\text{CCPL}}^{(r)} \mathcal{L}_{\text{CCPL}} + \lambda_{\text{EDL}} \mathcal{L}_{\text{EDL}}, \tag{20}$$

where λ_{PCL}, λ_{CCPL}, and λ_{EDL} denote the weights of the prototype contrastive loss, cross-client prototype contrastive loss, and evidential learning loss, respectively. $\mathcal{L}_{\text{PCL}}$ and $\mathcal{L}_{\text{CCPL}}$ are defined in Section 3.2.3, while $\mathcal{L}_{\text{EDL}}$ is defined in Section 3.4.1.

4 Experiments and Results

4.1 Experimental Settings

Datasets. We evaluate on PACS (4 domains, 7 classes) (Li et al., 2017), Office-Home (4 domains, 65 classes) (Venkateswara et al., 2017), Office-Caltech10 (4 domains, 10 classes) (Gong et al., 2012), and Terra Incognita (4 domains, 10 classes) (Beery et al., 2018). Following the leave-one-domain-out protocol, each source domain is split across two clients with a 90%/10% train/validation ratio; target domains are used entirely for testing.

Evaluation Metric. We report per-domain accuracy and the average across all target domains (Avg), following DomainBed (Gulrajani & Lopez-Paz, 2021).

Training Configuration. We train for 10 communication rounds, with 3 local epochs per round and a batch size of 32. SGD is used with an initial learning rate of 1×10^{-3} and weight decay of 5×10^{-4}. We adopt cosine learning-rate scheduling, with the classification head using a learning rate 10 times that of the backbone. Main experiments average over three seeds; ablation studies use a fixed seed.

4.2 Main Results

Tables 1 and 2 present classification accuracy of FedSP and existing methods on PACS, Office-Home, Office-Caltech10, and Terra Incognita. FedSP achieves the highest average accuracy on all four datasets: 86.45%, 75.49%, 96.88%, and 50.05%, outperforming second-best methods by 2.45,

Table 1: Main results on PACS and Office-Home.

Method	PACS					Office-Home				
	P	A	C	S	Avg	A	C	P	R	Avg
FedAvg (McMahan et al., 2017)	96.59	78.24	70.01	69.89	78.68	71.58	55.68	81.41	82.61	72.82
FedProx (Li et al., 2020)	96.83	77.99	72.14	75.45	80.60	72.04	57.40	82.01	83.32	73.69
FedIIR (Guo et al., 2023)	97.01	79.87	71.73	75.47	81.02	72.09	56.81	81.89	83.56	73.58
GA (Zhang et al., 2023b)	97.72	79.64	73.42	76.58	81.84	71.60	57.91	81.07	83.47	73.51
StableFDG (Park et al., 2023)	97.29	81.61	76.44	75.64	82.74	68.92	59.45	80.03	79.51	71.98
gPerXAN (Le et al., 2024)	94.61	81.61	**78.63**	76.32	82.79	68.45	55.90	79.06	81.25	71.17
FedLGF (Yan & Guo, 2024)	95.31	80.42	71.70	75.54	80.74	72.15	57.51	82.13	83.45	73.81
FedDIM (Liao et al., 2025)	96.85	80.55	73.24	76.42	81.76	72.71	57.76	82.03	83.24	73.94
DI (Zhou et al., 2025)	97.70	80.31	69.79	72.27	80.02	72.02	57.44	82.19	83.89	73.88
FedAlign (Gupta et al., 2025)	97.94	87.96	76.27	73.83	84.00	72.46	59.92	**83.41**	83.86	74.91
FedSP (Ours)	**98.26**	**91.03**	78.26	**78.24**	**86.45**	**73.20**	**61.26**	82.98	**84.50**	**75.49**

Table 2: Main results on Office-Caltech10 and Terra Incognita.

Method	Office-Caltech10					Terra Incognita				
	A	C	D	W	Avg	L_{100}	L_{38}	L_{43}	L_{46}	Avg
FedAvg(McMahan et al., 2017)	96.21	93.14	97.03	96.50	95.72	52.54	42.49	38.73	31.56	41.33
FedProx(Li et al., 2020)	96.10	92.79	97.24	97.18	95.83	58.48	43.06	40.07	32.43	43.51
FedIIR(Guo et al., 2023)	96.17	93.44	97.24	97.40	96.06	59.72	42.88	41.97	32.79	44.34
GA(Zhang et al., 2023b)	96.17	93.15	96.61	98.08	96.00	55.54	43.64	49.18	40.02	47.10
StableFDG(Park et al., 2023)	95.75	93.62	**99.36**	97.85	96.65	59.57	**44.38**	37.56	33.03	43.64
gPerXAN(Le et al., 2024)	95.02	91.16	98.73	96.95	95.46	56.61	30.68	46.75	30.04	41.02
FedLGF(Yan & Guo, 2024)	96.14	92.82	97.24	97.29	95.87	56.60	42.42	38.26	28.48	41.44
FedDIM(Liao et al., 2025)	96.24	93.02	97.66	97.40	96.08	59.72	43.16	40.37	33.77	44.26
DI(Zhou et al., 2025)	95.79	93.11	**99.36**	98.98	96.81	55.65	42.30	47.55	41.89	46.85
FedAlign(Gupta et al., 2025)	**96.35**	**95.04**	95.75	98.19	96.34	61.41	39.41	38.50	37.32	44.16
FedSP (Ours)	96.24	95.01	96.82	**99.43**	**96.88**	**65.72**	39.57	**49.44**	**45.48**	**50.05**

0.58, 0.54, and 5.89 percentage points, respectively. The improvements are particularly notable on PACS and Terra Incognita, which exhibit substantial cross-domain variations in visual styles and environmental conditions. These results demonstrate that jointly modeling shared knowledge and domain-specific discriminative information enhances adaptation to domain shifts and improves generalization on unseen domains.

4.3 Ablation Study

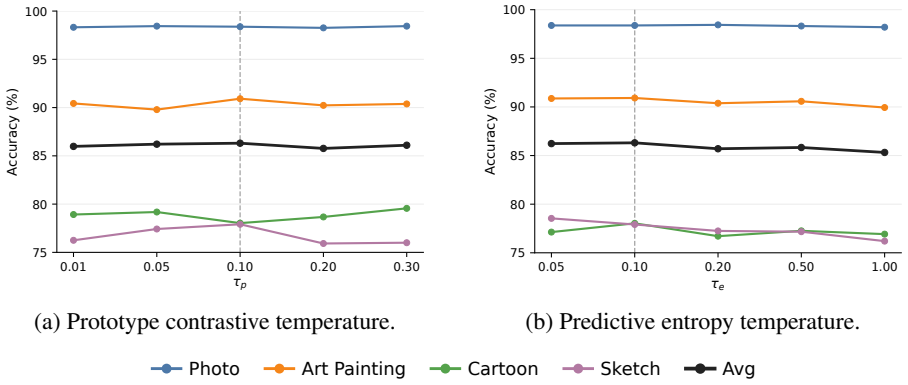
To evaluate the contribution of each module, we conduct ablation studies on PACS and Terra Incognita (Table 3). Adding CSAS, DSRA, or MUCL individually to the baseline improves average accuracy by 1.70, 0.28, and 1.30 points on PACS, and by 1.31, 0.73, and 3.47 points on Terra Incognita, respectively (Rows 1–3). Removing each module from the complete FedSP decreases accuracy by 1.35, 0.20, and 0.41 points on PACS, and by 1.83, 0.52, and 2.37 points on Terra Incognita, respectively (Rows 4–6).The complete model achieves the highest accuracy of 86.31% and 50.26% on both datasets, demonstrating the effectiveness and complementarity of all modules.

4.4 Hyperparameter Sensitivity Analysis

To evaluate the impact of key hyperparameters, we conduct sensitivity analyses of the prototype contrastive temperature τ_p and predictive entropy temperature τ_e on PACS (Figure 3). The model achieves the highest accuracy of 86.31% at $\tau_p = 0.1$, with accuracy ranging from 85.77% to 86.31% across settings, indicating stability to this parameter (Figure 3(a)). Similarly, the model peaks at 86.31% when $\tau_e = 0.1$ but drops to 85.32% at $\tau_e = 1.0$, as larger temperatures weaken entropy differences and reduce adaptive aggregation effectiveness (Figure 3(b)). We set both parameters to 0.1.

Table 3: Ablation results for component effectiveness. *w/ X* indicates that only module X is added to the FedAlign baseline, while *w/o X* indicates that the corresponding module is removed from the complete FedSP model.

Setting	PACS					Terra Incognita				
	P	A	C	S	Avg	L_{100}	L_{38}	L_{43}	L_{46}	Avg
w/ CSAS	97.90	89.50	79.01	76.00	85.60	65.59	43.87	39.79	34.78	46.01
w/ DSRA	97.84	88.13	76.02	74.73	84.18	62.68	41.17	41.53	36.33	45.43
w/ MUCL	98.44	89.60	76.88	75.90	85.20	61.86	40.52	46.98	43.32	48.17
w/o CSAS	98.26	89.89	74.83	76.86	84.96	62.43	40.66	44.96	45.68	48.43
w/o DSRA	98.56	90.14	79.10	76.66	86.11	65.20	42.34	48.47	42.95	49.74
w/o MUCL	98.08	89.40	78.37	77.73	85.90	67.00	41.82	43.79	38.94	47.89
FedAlign (Gupta et al., 2025)	97.72	87.99	76.15	73.73	83.90	62.21	42.19	37.02	37.38	44.70
FedSP (Ours)	98.38	90.92	78.03	77.91	**86.31**	67.02	40.63	48.42	44.96	**50.26**



(a) Prototype contrastive temperature.

(b) Predictive entropy temperature.

Figure 3: Hyperparameter sensitivity analysis on PACS.

5 Conclusion

We address the coupling between shared and domain-specific knowledge and differences in client reliability in federated domain generalization by proposing FedSP, a shared-personalized collaborative learning framework. FedSP learns transferable class semantics through category prototype-driven cross-domain alignment while preserving domain-specific discriminative attributes via residual adaptation, enabling complementary learning of shared and personalized knowledge. The proposed MUCL combines client-side evidential regularization with server-side predictive entropy-aware aggregation to improve local prediction reliability while dynamically adjusting client contributions. Experimental results demonstrate that exploiting the complementarity between class commonalities and domain-specific information, together with uncertainty-aware coordination, effectively alleviates the impact of data heterogeneity on model generalization. FedSP provides a new perspective for federated domain generalization under distributed heterogeneous data.

AI use statement

In this work, we used generative AI tools to assist with and polish the writing of the paper, retrieve and discover relevant literature, and support research ideation and execution.

We did not use generative AI tools to draft sections of the paper, generate synthetic datasets, or prove mathematical claims, and other required disclosure categories are not applicable to this work.

We have reviewed and verified all AI-assisted work. Specifically, AI-generated research ideas, methodological suggestions, and technical recommendations were independently analyzed, discussed, and experimentally validated by the authors. Relevant works identified with the assistance of AI tools were manually verified against the original publications. AI-assisted writing and language polishing were also manually reviewed to ensure their accuracy and consistency with the underlying

research. We take responsibility for the final content of this work, including text, research ideas, methodological descriptions, and other content produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

To ensure the reproducibility of our work, we provide comprehensive details across the paper, appendix, and supplementary materials. Section 4.1 describes the experimental settings, including datasets, evaluation metrics, and training configurations. Appendix A provides complete implementation details, including network architectures, hyperparameters, and computational environment specifications. All hyperparameters used in our experiments are explicitly listed, with no hidden configurations. The algorithm logic is formally described in Section 3. Additionally, we provide the main source code in the supplementary materials, including client-side training logic (client.py) and server-side aggregation (server.py). Our experiments follow the standard leave-one-domain-out protocol with three random seeds for main results and a fixed seed for ablation studies, ensuring fair comparison with prior work. All datasets used (PACS, Office-Home, Office-Caltech10, Terra Incognita) are publicly available, and we follow the standard data splits described in Section 4.1.

REFERENCES

Durmus Alp Emre Acar, Yue Zhao, Ramon Matas Navarro, Matthew Mattina, Paul N. Whatmough, and Venkatesh Saligrama. Federated learning based on dynamic regularization. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=B7v4QMR6Z9w.

Sara A. Al-Emadi, Yin Yang, and Ferda Ofli. Benchmarking object detectors under real-world distribution shifts in satellite imagery. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pp. 8299–8309. Computer Vision Foundation / IEEE, 2025. doi: 10.1109/CVPR52734.2025.00777. URL https://openaccess.thecvf.com/content/CVPR2025/html/Al-Emadi_Benchmarking_Object_Detectors_under_Real-World_Distribution_Shifts_in_Satellite_Imagery_CVPR_2025_paper.html.

Aristotelis Ballas and Christos Diou. Gradient-guided annealing for domain generalization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pp. 20558–20568. Computer Vision Foundation / IEEE, 2025. doi: 10.1109/CVPR52734.2025.01914. URL https://openaccess.thecvf.com/content/CVPR2025/html/Ballas_Gradient-Guided_Annealing_for_Domain_Generalization_CVPR_2025_paper.html.

Sara Beery, Grant Van Horn, and Pietro Perona. Recognition in terra incognita. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (eds.), *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part XVI*, volume 11220 of *Lecture Notes in Computer Science*, pp. 472–489. Springer, 2018. doi: 10.1007/978-3-030-01270-0_28. URL https://doi.org/10.1007/978-3-030-01270-0_28.

Gilles Blanchard, Gyemin Lee, and Clayton Scott. Generalizing from several related classification tasks to a new unlabeled sample. In John Shawe-Taylor, Richard S. Zemel, Peter L. Bartlett, Fernando C. N. Pereira, and Kilian Q. Weinberger (eds.), *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain*, pp. 2178–2186, 2011. URL https://proceedings.neurips.cc/paper/2011/hash/b571ecea16a9824023ee1af16897a582-Abstract.html.

Manh-Ha Bui, Toan Tran, Anh Tran, and Dinh Q. Phung. Exploiting domain-specific features to enhance domain generalization. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances*

in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual, pp. 21189–21201, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/b0f2ad44d26e1a6f244201fe0fd864d1-Abstract.html.

Junbum Cha, Sanghyuk Chun, Kyungjae Lee, Han-Cheol Cho, Seunghyun Park, Yunsung Lee, and Sungrae Park. SWAD: domain generalization by seeking flat minima. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 22405–22418, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/bcb41ccdc4363c6848a1d760f26c28a0-Abstract.html.

Hong-You Chen and Wei-Lun Chao. Fedbe: Making bayesian model ensemble applicable to federated learning. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=dgtpE6gKjHn.

Junming Chen, Meirui Jiang, Qi Dou, and Qifeng Chen. Federated domain generalization for image recognition via cross-client style transfer. In *IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2023, Waikoloa, HI, USA, January 2-7, 2023*, pp. 361–370. IEEE, 2023. doi: 10.1109/WACV56688.2023.00044. URL https://doi.org/10.1109/WACV56688.2023.00044.

Yiming Chen, Nan He, and Lifeng Sun. Fedawa: Aggregation weight adjustment in federated domain generalization. In *IEEE International Conference on Image Processing, ICIP 2024, Abu Dhabi, United Arab Emirates, october 27-30, 2024*, pp. 451–457. IEEE, 2024. doi: 10.1109/ICIP51287.2024.10647638. URL https://doi.org/10.1109/ICIP51287.2024.10647638.

Yu Ding, Lei Wang, Bin Liang, Shuming Liang, Yang Wang, and Fang Chen. Domain generalization by learning and removing domain-specific features. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/9941833e8327910ef25daeb9005e4748-Abstract-Conference.html.

Boqing Gong, Yuan Shi, Fei Sha, and Kristen Grauman. Geodesic flow kernel for unsupervised domain adaptation. In *2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012*, pp. 2066–2073. IEEE Computer Society, 2012. doi: 10.1109/CVPR.2012.6247911. URL https://doi.org/10.1109/CVPR.2012.6247911.

Ishaan Gulrajani and David Lopez-Paz. In search of lost domain generalization. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=lQdXeXDoWtI.

Yaming Guo, Kai Guo, Xiaofeng Cao, Tieru Wu, and Yi Chang. Out-of-distribution generalization of federated learning via implicit invariant relationships. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *International Conference on Machine Learning, ICML 2023, 23-29 July 2023, Honolulu, Hawaii, USA*, volume 202 of *Proceedings of Machine Learning Research*, pp. 11905–11933. PMLR, 2023. URL https://proceedings.mlr.press/v202/guo23b.html.

Sunny Gupta, Vinay Sutar, Varunav Singh, and Amit Sethi. Fedalign: Federated domain generalization with cross-client feature alignment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2025, Nashville, TN, USA, June 11-15, 2025*, pp. 1801–1810. Computer Vision Foundation / IEEE, 2025. URL https://openaccess.thecvf.com/content/CVPR2025W/FedVision/html/Gupta_FedAlign_Federated_Domain_Generalization_with_Cross-Client_Feature_Alignment_CVPRW_2025_paper.html.

Peter Kairouz, H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista A. Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, Rafael G. L. D'Oliveira, Hubert Eichner, Salim El Rouayheb, David Evans, Josh Gardner, Zachary Garrett, Adrià Gascón, Badih Ghazi, Phillip B. Gibbons, Marco Gruteser, Zaïd Harchaoui, Chaoyang He, Lie He, Zhouyuan Huo, Ben Hutchinson, Justin Hsu, Martin Jaggi, Tara Javidi, Gauri Joshi, Mikhail Khodak, Jakub Konečný, Aleksandra Korolova, Farinaz Koushanfar, Sanmi Koyejo, Tancrède Lepoint, Yang Liu, Prateek Mittal, Mehryar Mohri, Richard Nock, Ayfer Özgür, Rasmus Pagh, Hang Qi, Daniel Ramage, Ramesh Raskar, Mariana Raykova, Dawn Song, Weikang Song, Sebastian U. Stich, Ziteng Sun, Ananda Theertha Suresh, Florian Tramèr, Praneeth Vepakomma, Jianyu Wang, Li Xiong, Zheng Xu, Qiang Yang, Felix X. Yu, Han Yu, and Sen Zhao. Advances and open problems in federated learning. *Found. Trends Mach. Learn.*, 14(1-2):1–210, 2021. doi: 10.1561/2200000083. URL https://doi.org/10.1561/2200000083.

Daehee Kim, Youngjun Yoo, Seunghyun Park, Jinkyu Kim, and Jaekoo Lee. Selfreg: Self-supervised contrastive regularization for domain generalization. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, pp. 9599–9608. IEEE, 2021. doi: 10.1109/ICCV48922.2021.00948. URL https://doi.org/10.1109/ICCV48922.2021.00948.

Minyoung Kim, Da Li, and Timothy M. Hospedales. Domain generalisation via domain adaptation: An adversarial fourier amplitude approach. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=7IG0wsTND7w.

David Krueger, Ethan Caballero, Jörn-Henrik Jacobsen, Amy Zhang, Jonathan Binas, Dinghuai Zhang, Rémi Le Priol, and Aaron C. Courville. Out-of-distribution generalization via risk extrapolation (rex). In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, pp. 5815–5826. PMLR, 2021. URL http://proceedings.mlr.press/v139/krueger21a.html.

Khiem Le, Long Ho, Cuong Do, Danh Le Phuoc, and Kok-Seng Wong. Efficiently assemble normalization layers and regularization for federated domain generalization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024*, pp. 6027–6036. IEEE, 2024. doi: 10.1109/CVPR52733.2024.00576. URL https://doi.org/10.1109/CVPR52733.2024.00576.

Da Li, Yongxin Yang, Yi-Zhe Song, and Timothy M. Hospedales. Deeper, broader and artier domain generalization. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pp. 5543–5551. IEEE Computer Society, 2017. doi: 10.1109/ICCV.2017.591. URL https://doi.org/10.1109/ICCV.2017.591.

Fangyu Li, Xuqiang Chen, Zhu Han, Yongping Du, and Honggui Han. Federated learning via reweighting information bottleneck with domain generalization. *Inf. Sci.*, 677:120825, 2024. doi: 10.1016/J.INS.2024.120825. URL https://doi.org/10.1016/j.ins.2024.120825.

Jiansheng Li, Xingxuan Zhang, Hao Zou, Yige Guo, Renzhe Xu, Yilong Liu, Chuzhao Zhu, Yue He, and Peng Cui. COUNTS: benchmarking object detectors and multimodal large language models under distribution shifts. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, pp. 9186–9198. Computer Vision Foundation / IEEE, 2025a. doi: 10.1109/CVPR52734.2025.00858. URL https://openaccess.thecvf.com/content/CVPR2025/html/Li_COUNTS_Benchmarking_Object_Detectors_and_Multimodal_Large_Language_Models_under_CVPR_2025_paper.html.

Pan Li, Da Li, Wei Li, Shaogang Gong, Yanwei Fu, and Timothy M. Hospedales. A simple feature augmentation for domain generalization. In *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*, pp. 8866–8875. IEEE, 2021a. doi: 10.1109/ICCV48922.2021.00876. URL https://doi.org/10.1109/ICCV48922.2021.00876.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. In Inderjit S. Dhillon, Dimitris S. Papailiopoulos, and Vivienne Sze (eds.), *Proceedings of the Third Conference on Machine Learning and Systems, MLSys 2020, Austin, TX, USA, March 2-4, 2020*. mlsys.org, 2020. URL https://proceedings.mlsys.org/paper_files/paper/2020/hash/1f5fe83998a09396ebe6477d9475ba0c-Abstract.html.

Xiaoxiao Li, Meirui Jiang, Xiaofei Zhang, Michael Kamp, and Qi Dou. Fedbn: Federated learning on non-iid features via local batch normalization. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021b. URL https://openreview.net/forum?id=6YEQUn0QICG.

Ying Li, Xingwei Wang, Rongfei Zeng, Praveen Kumar Donta, Ilir Murturi, Min Huang, and Schahram Dustdar. Federated domain generalization: A survey. *Proc. IEEE*, 113(4):370–410, 2025b. doi: 10.1109/JPROC.2025.3596173. URL https://doi.org/10.1109/JPROC.2025.3596173.

Tianchi Liao, Binghui Xie, Lele Fu, Sheng Huang, Bowen Deng, Chuan Chen, and Zibin Zheng. Federated domain generalization with decision insight matrix. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2025, Montreal, Canada, August 16-22, 2025*, pp. 5689–5697. ijcai.org, 2025. doi: 10.24963/IJCAI.2025/633. URL https://doi.org/10.24963/ijcai.2025/633.

Quande Liu, Cheng Chen, Jing Qin, Qi Dou, and Pheng-Ann Heng. Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*, pp. 1013–1023. Computer Vision Foundation / IEEE, 2021. doi: 10.1109/CVPR46437.2021.00107. URL https://openaccess.thecvf.com/content/CVPR2021/html/Liu_FedDG_Federated_Domain_Generalization_on_Medical_Image_Segmentation_via_Episodic_CVPR_2021_paper.html.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In Aarti Singh and Xiaojin (Jerry) Zhu (eds.), *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, 20-22 April 2017, Fort Lauderdale, FL, USA*, volume 54 of *Proceedings of Machine Learning Research*, pp. 1273–1282. PMLR, 2017. URL http://proceedings.mlr.press/v54/mcmahan17a.html.

Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *Proceedings of the 30th International Conference on Machine Learning, ICML 2013, Atlanta, GA, USA, 16-21 June 2013*, volume 28 of *JMLR Workshop and Conference Proceedings*, pp. 10–18. JMLR.org, 2013. URL http://proceedings.mlr.press/v28/muandet13.html.

Trong-Binh Nguyen, Duong Minh Nguyen, Jinsun Park, Viet Quoc Pham, and Won-Joo Hwang. Federated domain generalization with data-free on-server matching gradient. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*. OpenReview.net, 2025. URL https://openreview.net/forum?id=8TERgu1Lb2.

Jungwuk Park, Dong-Jun Han, Jinho Kim, Shiqiang Wang, Christopher G. Brinton, and Jaekyun Moon. Stablefdg: Style and attention based learning for federated domain generalization. In Alice Oh, Tristan Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/dae8bdacd265399b193e6b43d44a80f0-Abstract-Conference.html.

Farhad Pourpanah, Mahdiyar Molahasani, Milad Soltany, Michael A. Greenspan, and Ali Etemad. Federated unsupervised domain generalization using global and local alignment of gradients. In Toby Walsh, Julie Shah, and Zico Kolter (eds.), *Thirty-Ninth AAAI Conference on Artificial Intelligence, Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence, Fifteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2025, Philadelphia, PA,*

USA, February 25 - March 4, 2025, pp. 19948–19958. AAAI Press, 2025. doi: 10.1609/AAAI.V39I19.34197. URL https://doi.org/10.1609/aaai.v39i19.34197.

Alexandre Ramé, Corentin Dancette, and Matthieu Cord. Fishr: Invariant gradient variances for out-of-distribution generalization. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvári, Gang Niu, and Sivan Sabato (eds.), *International Conference on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA*, volume 162 of *Proceedings of Machine Learning Research*, pp. 18347–18377. PMLR, 2022. URL https://proceedings.mlr.press/v162/rame22a.html.

Ozan Sener and Vladlen Koltun. Domain generalization without excess empirical risk. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/57568e093cbe0a222de0334b36e83cf5-Abstract-Conference.html.

Murat Sensoy, Lance M. Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pp. 3183–3193, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/a981f2b708044d6fb4a71a1463242520-Abstract.html.

Yue Tan, Guodong Long, Lu Liu, Tianyi Zhou, Qinghua Lu, Jing Jiang, and Chengqi Zhang. Fedproto: Federated prototype learning across heterogeneous clients. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*, pp. 8432–8440. AAAI Press, 2022. doi: 10.1609/AAAI.V36I8.20819. URL https://doi.org/10.1609/aaai.v36i8.20819.

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 2017, Honolulu, HI, USA, July 21-26, 2017*, pp. 5385–5394. IEEE Computer Society, 2017. doi: 10.1109/CVPR.2017.572. URL https://doi.org/10.1109/CVPR.2017.572.

Chenyang Wang, Junjun Jiang, Xingyu Hu, Xianming Liu, and Xiangyang Ji. Federated domain generalization via data-centric flatness optimization. *Pattern Recognit.*, 174:113023, 2026. doi: 10.1016/J.PATCOG.2025.113023. URL https://doi.org/10.1016/j.patcog.2025.113023.

Zhe Wang, Jake Grigsby, and Yanjun Qi. Pgrad: Learning principal gradients for domain generalization. In *The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023*. OpenReview.net, 2023. URL https://openreview.net/forum?id=CgCmwcfgEdH.

Jianhan Wu, Xiaoyang Qu, Zhangcheng Huang, and Jianzong Wang. Federated domain generalization with domain-specific soft prompts generation. In *IEEE/CVF International Conference on Computer Vision, ICCV 2025, Honolulu, HI, USA, October 19-25, 2025*, pp. 2366–2375. IEEE, 2025. doi: 10.1109/ICCV51701.2025.00228. URL https://doi.org/10.1109/ICCV51701.2025.00228.

Hao Yan and Yuhong Guo. Local and global flatness for federated domain generalization. In Ales Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (eds.), *Computer Vision - ECCV 2024 - 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part LXXXIII*, volume 15141 of *Lecture Notes in Computer Science*, pp. 71–87. Springer, 2024. doi: 10.1007/978-3-031-73010-8_5. URL https://doi.org/10.1007/978-3-031-73010-8_5.

Fengxiang Yang, Zhun Zhong, Zhiming Luo, Yifan He, Shaozi Li, and Nicu Sebe. Diversity-authenticity co-constrained stylization for federated domain generalization in person re-identification. In Michael J. Wooldridge, Jennifer G. Dy, and Sriraam Natarajan (eds.), *Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2024, February 20-27, 2024, Vancouver, Canada*, pp. 6477–6485. AAAI Press, 2024a. doi: 10.1609/AAAI.V38I6.28468. URL https://doi.org/10.1609/aaai.v38i6.28468.

Seunghan Yang, Seokeon Choi, Hyunsin Park, Sungha Choi, Simyung Chang, and Sungrack Yun. Feature diversification and adaptation for federated domain generalization. In Ales Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (eds.), *Computer Vision - ECCV 2024 - 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part LXXII*, volume 15130 of *Lecture Notes in Computer Science*, pp. 52–70. Springer, 2024b. doi: 10.1007/978-3-031-73220-1_4. URL https://doi.org/10.1007/978-3-031-73220-1_4.

Liling Zhang, Xinyu Lei, Yichun Shi, Hongyu Huang, and Chao Chen. Federated learning for iot devices with domain generalization. *IEEE Internet Things J.*, 10(11):9622–9633, 2023a. doi: 10.1109/JIOT.2023.3234977. URL https://doi.org/10.1109/JIOT.2023.3234977.

Ruipeng Zhang, Qinwei Xu, Jiangchao Yao, Ya Zhang, Qi Tian, and Yanfeng Wang. Federated domain generalization with generalization adjustment. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, pp. 3954–3963. IEEE, 2023b. doi: 10.1109/CVPR52729.2023.00385. URL https://doi.org/10.1109/CVPR52729.2023.00385.

Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. URL https://openreview.net/forum?id=6xHJ37MVxxp.

Kaiyang Zhou, Ziwei Liu, Yu Qiao, Tao Xiang, and Chen Change Loy. Domain generalization: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(4):4396–4415, 2023. doi: 10.1109/TPAMI.2022.3195549. URL https://doi.org/10.1109/TPAMI.2022.3195549.

Zikun Zhou, Wen Huang, Xingyi Wang, Zhishuo Zhang, Zhun Zhang, Jian Peng, and Feihu Huang. Improving federated domain generalization through dynamical weights calculated from data influences on global model update. In Toby Walsh, Julie Shah, and Zico Kolter (eds.), *Thirty-Ninth AAAI Conference on Artificial Intelligence, Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence, Fifteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2025, Philadelphia, PA, USA, February 25 - March 4, 2025*, pp. 23045–23053. AAAI Press, 2025. doi: 10.1609/AAAI.V39I21.34468. URL https://doi.org/10.1609/aaai.v39i21.34468.

A Appendix

A.1 Implementation Details

A.1.1 Network Architecture

We use an ImageNet-pretrained ResNet-50 as the feature extractor and configure the classification head according to the number of classes in each dataset. The input images are uniformly resized to 224×224 and normalized using the ImageNet mean and standard deviation. The output feature dimension is 2048.

A.1.2 Hyperparameters

Table 4 summarizes all hyperparameters used in FedSP.

Table 4: Hyperparameters for FedSP.

Parameter	Symbol	Value
Base Training		
Cross-entropy loss weight	–	1.0
Supervised contrastive loss weight	λ_1	0.1
JS divergence loss weight	λ_2	0.1
Consistency loss weight	λ_3	0.1
CSAS Module		
Prototype contrastive loss weight	λ_{PCL}	0.5
Cross-client prototype loss weight	λ_{CCPL}	0.5
Prototype contrastive temperature	τ_p	0.1
Cross-client contrastive temperature	τ_c	0.3
Multi-view aggregation weight	λ	0.7
Prototype momentum coefficient	η_p	0.9
Aggregation temperature	τ_a	0.5
Shrinkage coefficient	ζ	0.3
Smoothing term	ε	1×10^{-6}
Warm-up rounds	–	1
Ramp-up rounds	–	2
DSRA Module		
DSRA learning rate	–	1×10^{-3}
DSRA weight decay	–	1×10^{-4}
DSRA hidden dimension	–	64
DSRA momentum coefficient	η_ψ	0.7
Residual strength initial value	α	sigmoid(-1.0)$\times$0.5
MUCL Module		
Evidential loss weight	λ_{EDL}	0.1
KL regularization weight	λ_{KL}	1.0
EDL head learning rate	–	1×10^{-3}
EDL head weight decay	–	1×10^{-4}
EDL annealing rounds	–	3
Predictive entropy temperature	τ_e	0.1
Fusion coefficient	λ_e	0.5

A.1.3 Computational Environment

All experiments are conducted on a single NVIDIA GeForce RTX 5090 GPU (32GB memory) with CUDA 13.0. The implementation is based on PyTorch with the corresponding CUDA toolkit. For multi-domain leave-one-domain-out experiments, we run each target domain sequentially or in parallel depending on memory availability.