FedVR: Variance Regularized Hypernetwork for Federated Domain Generalization

Sunny Gupta, Shambhavi Shanker, Amit Sethi
Indian Institute of Technology Bombay, India
{sunnygupta, 21d070066, asethi}@iitb.ac.in

Abstract

Federated Learning (FL) enables model training across decentralized and privacy-sensitive data sources, but its effectiveness is severely degraded by domain shifts. Conventional domain generalization methods aim to extract invariant features, yet their integration into FL is hindered by privacy constraints and the limitations of linear aggregation in FedAvg. Hypernetwork-based approaches offer non-linear parameter synthesis, but existing methods lack explicit fairness guarantees and remain fragile under strong heterogeneity. We propose FedVR, a framework for hypernetwork-driven generalization adjustment in federated domain generalization. FedVR generates per-domain models via a hypernetwork conditioned on distributional embeddings that summarize each client's data statistics. To improve robustness, we extend generalization adjustment to these hypernetwork-generated models, explicitly minimizing the variance of per-domain generalization gaps. Moreover, FedVR enables zero-shot deployment on unseen domains: given a small unlabeled or lightly labeled pilot set, the domain encoder produces a new embedding from which the hypernetwork synthesizes specialized parameters without fine-tuning. Extensive experiments on PACS, Office-Home, and VLCS benchmarks demonstrate that FedVR achieves superior accuracy, fairness, and calibration across both in-domain and out-of-domain settings, outperforming strong federated baselines.

Introduction

Deep neural networks (DNNs) have achieved remarkable success, primarily relying on large, centralized datasets (Krizhevsky, Sutskever, and Hinton 2012; He et al. 2016; Dosovitskiy et al. 2020; Liu et al. 2021b). However, in several real-world scenarios, data is often fragmented across silos due to privacy rules that forbid sharing. Federated learning (FL) (McMahan et al. 2017; Li et al. 2020) addresses this by enabling collaborative model training without raw data exchange, but faces two core obstacles: (i) non-i.i.d. client distributions which hinder model convergence and performance (Li et al. 2019; Karimireddy et al. 2020), and (ii) domain shift, where the trained model must generalize effectively to systematically different, unseen test data (Liu et al. 2021a; Zhang et al. 2023b; Bai, Bagchi, and Inouye 2023). This motivates the demanding problem of *federated domain generalization* (FDG) training a model across decentralized heterogeneous source domains to robustly generalize to *unseen target domains*, while adhering to strict communication and privacy constraints (Zhang et al. 2021; Yuan et al. 2023; Seokeon et al. 2024).

Most FDG methods rely on *parameter-space fusion*(e.g., FedAvg and its variants (McMahan et al. 2017; Li et al. 2019; Karimireddy et al. 2020; Wang et al. 2020)), linearly aggregating client-specific models or adapters. While simple, this approach is constrained by a *convex-fusion ceiling*: the global model can only interpolate within the convex hull of source parameters, leaving unseen domains poorly represented (Zhang et al. 2023b; Yuan et al. 2023; Bai, Bagchi, and Inouye 2023). Moreover, transmitting large parameter blocks inflates bandwidth requirements and heightens privacy risk (Geiping et al. 2020; Huang et al. 2021; Hatamizadeh et al. 2023).

To overcome these limitations, we introduce a new paradigm: Variance-Regularized Hypernetwork Generalization (FedVR) for federated domain generalization. Our framework shifts away from conventional parameter averaging and gradient transmission, and instead learns a *hypernetwork* that dynamically generates client-specific model parameters conditioned on domain embeddings. This hypernetwork-driven personalization provides flexible adaptation to heterogeneous local data while maintaining a unified global structure. To ensure stable convergence amidst cross-domain heterogeneity, we integrate two novel schemes: (i) *Variance-Regularized Aggregation*, which explicitly penalizes inter-client generalization disparity, promoting fairness and stable convergence, and (ii) *Generalization-Weighted Optimization*, where the global server rebalances client contributions based on their generalization consistency over training rounds, enhancing robustness against noisy or overly-dominant domains.

Key Contributions

- **Hypernetwork-driven personalization:** A domain-conditioned hypernetwork to synthesize client-specific parameters, enabling secure flexible adaptation.
- **Variance-regularization:** An explicit inter-domain variance penalty to ensure stability and fairness across heterogeneous clients.

- **Generalization-weighted optimization:** Client updates are reweighted based on their generalization consistency rather than empirical performance.

FedVR framework thus provides a unified mechanism that bridges *federated personalization* and *domain generalization*, resulting in a globally coherent yet client-adaptive model that capable of strong, zero-shot generalization to unseen target domains.

Related Work

Federated Domain Generalization

Federated Domain Generalization (FDG) combines the privacy and heterogeneity challenges of FL with the distribution gap inherent in domain generalization (DG) (Li et al. 2018; McMahan et al. 2017; Bai, Bagchi, and Inouye 2023). It aims to train across decentralized source domains to robustly generalize to unseen target domains. Current FDG literature employs diverse strategies such as alignment and correlation Mitigation where methods focus on standardizing representations or updates, and adaptive and disentangled parameterization, where adaptation mechanisms are explored. FedAlign (Gupta et al. 2025) introduced cross-client feature alignment to enforce representation consistency across clients, a goal shared by FedVR but achieved implicitly via variance regularization. Pourpanah et al. (2025) introduced gradient alignment in unsupervised FDG settings, aiming to harmonize local update directions. Other approaches, such as Ma et al. (2024), tackle FDG by focusing on reducing spurious correlations, which can impair generalization to novel distributions. These contrast with our approach, which stabilizes generalization using variance-regularized aggregation of hypernetwork parameters, avoiding explicit sharing of features or local gradients entirely. Considering adaptive methods, Zhang et al. (2023a) proposed disentangling domain-specific factors and aggregating domain experts, a paradigm conceptually related to the parameterized weights generated by our hypernetwork.

In summary, existing FDG techniques often rely on explicit feature or gradient sharing for alignment, or employ hard-aggregation of domain-specific components (e.g., adapters or experts). Our FedVR framework fundamentally shifts the communication primitive to the hypernetwork's small, shared parameters, mitigating privacy leakage and the convex-fusion ceiling while achieving cross-domain stability through implicit, variance-based regularization.

Hypernetworks for FL

Hypernetworks dynamically generate the weights of a larger target network, significantly reducing communication and storage overhead in FL (Ha, Dai, and Le 2016). This generative nature also weakens the direct link between shared parameters and local data, providing an intrinsic privacy benefit (Carey, Du, and Wu 2022). In hypernetwork-based FL, a server-side generator H_ϕ maps a compact client embedding e_k to client parameters $\theta_k = H_\phi(e_k)$ enabling interpolation in the embedding space (Ha, Dai, and Le 2016; Shamsian et al. 2021; Carey, Du, and Wu 2022; Li et al. 2023; Tashakori et al. 2023; Lin et al. 2023). This induces a bi-level inversion problem (over both H_ϕ and e_k), thereby reducing potential privacy leakage. Beyond FL, hypernetworks have shown strong promise in multi-domain and multi-task learning (Rosenbaum, Klinger, and Riemer 2017; Kumar et al. 2022), where they enable joint learning by sharing knowledge across related tasks. For instance, HMOE (Qu et al. 2022) introduced a hypernetwork-based Mixture-of-Experts (MoE) for domain generalization that learns an embedding space and minimizes divergence between predicted and ground-truth embeddings.

Within FL, however, this line of research remains relatively nascent. Shamsian et al. (Shamsian et al. 2021) introduced pFedHN, the first hypernetwork-based alternative to FedAvg, which generates client-specific parameters asynchronously from embeddings. Ma et al. (Ma et al. 2022) further extended this idea to layer-wise parameter generation, modeling mutual client contributions. Recently, Bartholet et al. (Bartholet et al. 2024) proposed hFedF, a hypernetwork-driven model fusion framework that addresses stability and convergence issues in federated training. Yet, most of these works primarily target either *personalization* or server-side fusion, and thus fall short of addressing the broader challenge of *FDG*. In particular, existing approaches do not clarify how to *compose* knowledge across multiple source domains to represent an unseen test domain, nor how to couple generator-driven parameterization with vision-specific MoE specialization, both of which are crucial for robust FDG. (Li et al. 2025) proposed a hypernetwork aggregation mechanism for decentralized personalized FL, achieving efficient client adaptation. (Zhou et al. 2025) extended this concept to enable model deployment to non-participating clients without fine-tuning via hypernetwork inference. HyperFedNet (Chen et al. 2024) further demonstrated that hypernetwork-based client synthesis significantly reduces communication cost. FedVR generalizes these settings by introducing variance-harmonized aggregation that promotes fairness and stability while supporting zero-shot parameter generation for unseen domains.

More recently, parameter-efficient vision methods have been explored. FedDG-MoE (Radwan et al. 2025) instantiated a frozen pre-trained ViT with client-specialized MoE adapters and a test-time statistical fusion rule. Specifically, cosine similarity between test features and client-tracked moments determined adapter weights, effectively implementing *parameter-space convex averaging*. While effective, this paradigm suffers from three limitations: (i) a *convex-fusion ceiling*, since linear averaging cannot capture the non-linear structure of unseen domains; (ii) high *per-client storage and communication* overhead due to maintaining distinct adapters; and (iii) an enlarged *privacy surface*, since transmitting rich client statistics exposes sensitive distributional information (Guo et al. 2025).

Our FedVR framework addresses this gap by coupling hypernetwork-driven parameter synthesis, which inherently moves beyond convex interpolation with explicit variance-regularized aggregation to promote stable and robust generalization across heterogeneous domains.

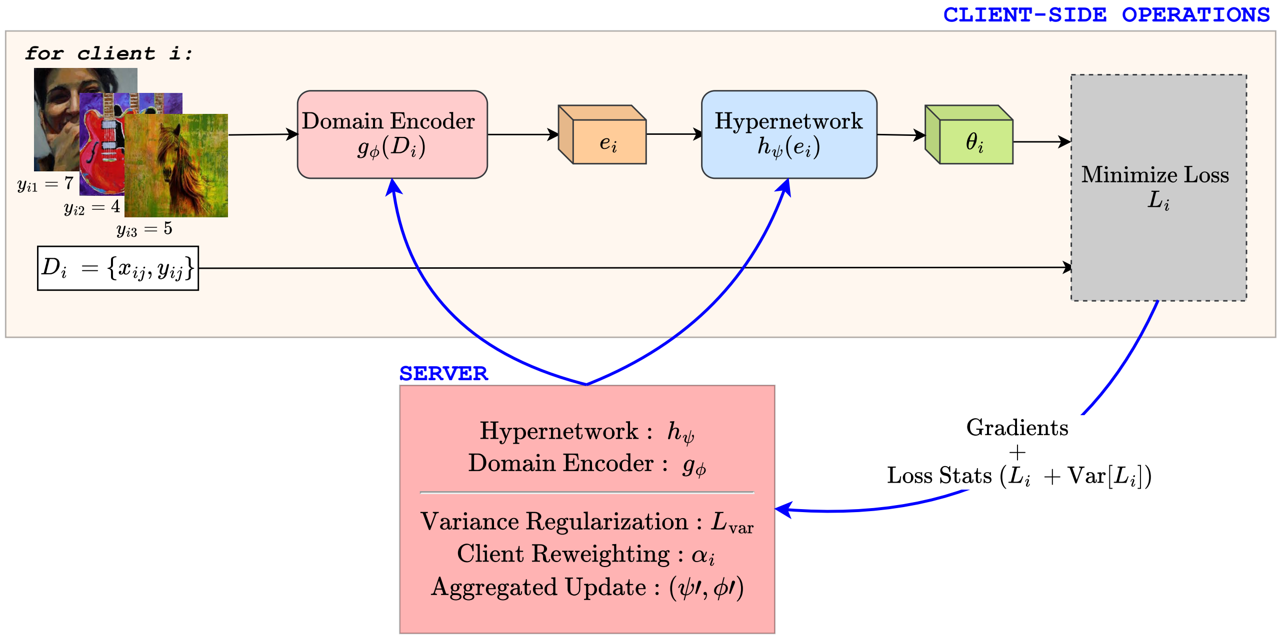

Figure 1: **FedVR** framework. Each client domain D_i is encoded into an embedding $e_i = g_\phi(D_i)$, which the hypernetwork $h_\psi(e_i)$ maps to domain-specific parameters θ_i. Clients perform local optimization on (θ_i, ϕ) and send gradients along with loss statistics $(L_i, \mathrm{Var}[L_i])$ to the server. The server updates global parameters $\{\psi, \phi\}$ via fairness-aware aggregation with variance regularization (L_{var}) and reweighting coefficients (α_i). For unseen domains D_{new}, FedVR generates parameters $\theta_{\text{new}} = h_\psi(g_\phi(D_{\text{new}}))$ for zero-shot inference.

Methodology

To address this challenge, we propose a Variance-Regularized Hypernetwork Generalization framework that unifies three principles: (1) adaptive client modeling through a domain-conditioned hypernetwork, (2) variance-based regularization to align generalization behavior across clients, and (3) dynamic generalization weighting for stable aggregation.

FedVR, bridges federated personalization and domain generalization by learning variance-aware hypernetworks that adaptively synthesize client-specific model parameters while preserving global coherence.

Framework Overview

Consider a federation of N clients, each holding a private dataset $\mathcal{D}_i = \{(x_i, y_i)\}$ sampled from distribution $p_i(x, y)$. The goal is to train a shared model that generalizes across clients without sharing raw data. The global objective is:

$$\min_\theta \; \mathbb{E}_{i\sim\mathcal{C}}\Big[\mathcal{L}_i(f_{\theta_i}(x_i), y_i)\Big], \tag{1}$$

where f_{θ_i} represents the model at client i, parameterized by θ_i that varies across domains.

Instead of learning θ_i directly, we parameterize them using a *domain-conditioned hypernetwork* that generates weights conditioned on client-specific embeddings. To make optimization consistent between clients and the server, we explicitly separate local and global gradient computations. Each client transmits $\{\nabla_\psi\mathcal{L}_i, \nabla_\phi\mathcal{L}_i, \bar{\mathcal{L}}_i, \mathrm{Var}[\mathcal{L}_i]\}$, while the server constructs the inter-client variance term $\mathcal{L}_{\text{var}}$ and its gradient $\nabla_\psi\mathcal{L}_{\text{var}}$ for regularized aggregation. In implementation, the local variance term $\mathrm{Var}[\mathcal{L}_i]$ is estimated across mini-batch losses within each communication round:

$$\mathrm{Var}[\mathcal{L}_i] = \mathrm{Var}_{b\in\mathcal{B}_i}[\mathcal{L}_i^b], \tag{2}$$

where $\mathcal{B}_i$ denotes the set of batches processed by client i. This per-round formulation captures short-term instability in local optimization without requiring cross-round memory. This resolves the gradient dependency mismatch between client-side updates and server aggregation steps.

Domain Encoder and Local Variance

Each client i first computes a domain representation using the encoder g_ϕ. Specifically, the encoder operates on aggregated feature statistics derived from the local dataset:

$$e_i = g_\phi(\bar{x}_i), \quad \bar{x}_i = \frac{1}{|\mathcal{D}_i|}\sum_{(x,y)\in\mathcal{D}_i} f_{\text{backbone}}(x),$$

where $f_{\text{backbone}}(\cdot)$ extracts intermediate feature representations. The resulting embedding e_i captures domain-level characteristics such as texture bias, color distribution, or class co-occurrence patterns. Both g_ϕ and the hypernetwork h_ψ are jointly optimized through the global aggregation rule in Algorithm 1.

For each client, the local loss variance $\mathrm{Var}[\mathcal{L}_i]$ is computed across mini-batches within a communication round:

$$\mathrm{Var}[\mathcal{L}_i] = \frac{1}{B}\sum_{b=1}^{B}(\mathcal{L}_{i,b} - \bar{\mathcal{L}}_i)^2, \quad \bar{\mathcal{L}}_i = \frac{1}{B}\sum_{b=1}^{B}\mathcal{L}_{i,b},$$

where $\mathcal{L}_{i,b}$ is the mini-batch loss and B denotes the number of local batches. This variance serves as a measure of optimization stability and generalization reliability for client i.

Hypernetwork-Based Client Adaptation

For each client i, we define a hypernetwork h_ψ that maps a domain embedding e_i to client-specific model parameters:

$$\theta_i = h_\psi(e_i), \quad e_i = g_\phi(\mathcal{D}_i), \tag{3}$$

where g_ϕ denotes a domain encoder that extracts a statistical representation of the local data distribution. Unless stated otherwise, h_ψ generates the classifier head and adapter parameters of f_{θ_i}, while the shared convolutional backbone is frozen to minimize communication and improve stability.

The hypernetwork h_ψ thus learns a continuous function over the space of domain embeddings, generating personalized weights θ_i that respect the inter-domain geometry.

Unlike standard aggregation (e.g., FedAvg), the server does not average model parameters directly. Instead, it updates ψ to minimize the global empirical risk:

$$\min_{\psi,\phi}\ \mathbb{E}_i\Big[\mathcal{L}_i(f_{h_\psi(g_\phi(\mathcal{D}_i))};\mathcal{D}_i)\Big]. \tag{4}$$

This formulation enables task-specific adaptation via h_ψ while retaining a shared global knowledge space.

Domain encoder g_ϕ specification: In all experiments, g_ϕ is implemented as a lightweight two-layer multilayer perceptron (MLP) that maps the feature statistics of each client's local dataset to a 128-dimensional embedding e_i. For vision tasks, g_ϕ operates on mean-pooled penultimate features extracted from the backbone encoder f_{θ_i}, computed over the client's local dataset (unlabeled samples suffice). The resulting embedding captures global domain-level statistics such as style, feature dispersion, and activation variance. Embeddings e_i are recomputed once per communication round to summarize updated domain information while avoiding batch-level noise.

Algorithm 1: **FedVR: Variance-Harmonized Hypernetwork for Federated Domain Generalization**

Require: N clients with datasets $\{\mathcal{D}_i\}_{i=1}^N$, hypernetwork h_ψ, domain encoder g_ϕ, learning rate η, variance weight λ_{var}, temperature τ, and total communication rounds T.

1: **Server initialization:** Initialize (ψ^0, ϕ^0) and broadcast to all clients.
2: **for** $t = 1$ to T **do**
Client-side computation (in parallel for each i):
3: $e_i \leftarrow g_{\phi^t}(\mathcal{D}_i)$ // Domain embedding
4: $\theta_i \leftarrow h_{\psi^t}(e_i)$ // Personalized parameters
5: Train local model f_{θ_i} on $\mathcal{D}_i$ to minimize $\mathcal{L}_i$
6: Compute mean $\bar{\mathcal{L}}_i$ and variance $\mathrm{Var}[\mathcal{L}_i]$
7: Send $\{\nabla_\psi\mathcal{L}_i, \nabla_\phi\mathcal{L}_i, \bar{\mathcal{L}}_i, \mathrm{Var}[\mathcal{L}_i]\}$ to server
Server-side aggregation:
8: Compute inter-client loss variance:

$$\mathcal{L}_{\text{var}} = \frac{1}{N}\sum_{i=1}^{N}(\bar{\mathcal{L}}_i - \bar{\mathcal{L}})^2, \quad \bar{\mathcal{L}} = \frac{1}{N}\sum_{i=1}^{N}\bar{\mathcal{L}}_i$$

9: Compute generalization weights:

$$\alpha_i = \frac{\exp(-\tau\,\mathrm{Var}[\mathcal{L}_i])}{\sum_{j=1}^{N}\exp(-\tau\,\mathrm{Var}[\mathcal{L}_j])}$$

10: Compute variance-regularized gradient:

$$\nabla_\psi\mathcal{L}_{\text{var}} = \frac{2}{N}\sum_{i=1}^{N}(\bar{\mathcal{L}}_i - \bar{\mathcal{L}})\,\nabla_\psi\mathcal{L}_i$$

11: Aggregate gradients with regularization:

$$\nabla_\psi^{\text{global}} = \sum_{i=1}^{N}\alpha_i\nabla_\psi\mathcal{L}_i + \lambda_{\text{var}}\,\nabla_\psi\mathcal{L}_{\text{var}},$$
$$\nabla_\phi^{\text{global}} = \sum_{i=1}^{N}\alpha_i\nabla_\phi\mathcal{L}_i$$

12: Update parameters: $\psi^{t+1} \leftarrow \psi^t - \eta\nabla_\psi^{\text{global}}$, $\phi^{t+1} \leftarrow \phi^t - \eta\nabla_\phi^{\text{global}}$
13: Broadcast (ψ^{t+1}, ϕ^{t+1}) to all clients
14: **end for**
15: **Output:** Trained hypernetwork parameters ψ^T generating domain-adaptive weights $\theta_i = h_{\psi^T}(g_{\phi^T}(\mathcal{D}_i))$ for unseen domains.

Note: In Step 10, the server computes $\nabla_\psi\mathcal{L}_{\text{var}} = \frac{2}{N}\sum_i(\bar{\mathcal{L}}_i - \bar{\mathcal{L}})\nabla_\psi\mathcal{L}_i$ using only client-provided summaries, without access to raw data.

Variance-Regularized Generalization

Although hypernetworks enable client-specific adaptation, domain heterogeneity can still lead to imbalanced generalization across clients. To mitigate this, we introduce a variance-regularized objective that penalizes large inter-client disparities in the mean local losses. For clarity, we denote by $\mathcal{L}_i$ the average empirical loss of client i over its local dataset $\mathcal{D}_i$, and by $\bar{\mathcal{L}}_i$ the mini-batch mean loss within a communication round. The global variance regularizer is defined as

$$\mathcal{L}_{\text{var}} = \frac{1}{N}\sum_{i=1}^{N}(\mathcal{L}_i - \bar{\mathcal{L}})^2, \qquad \bar{\mathcal{L}} = \frac{1}{N}\sum_{i=1}^{N}\mathcal{L}_i. \tag{5}$$

This term enforces cross-domain consistency, discouraging any client from dominating or diverging due to distributional shifts. The complete server-side training objective becomes:

$$\min_{\psi,\,\phi}\ \mathbb{E}_i\Big[\,\mathcal{L}_i(f_{h_\psi(e_i)};\mathcal{D}_i)\,\Big] + \lambda_{\text{var}}\,\mathcal{L}_{\text{var}}, \tag{6}$$

where λ_{var} controls the regularization strength.

Notation clarification: We use $\mathcal{L}_i$ consistently to denote the mean loss of client i. The domain embedding e_i is computed as $e_i = g_\phi(\bar{x}_i)$, where $\bar{x}_i = \frac{1}{|\mathcal{D}_i|}\sum_{(x,y)\in\mathcal{D}_i} f_{\text{backbone}}(x)$ represents the aggregated feature statistics of $\mathcal{D}_i$. This ensures that both Section 3.2 and Figure 1 refer to the same input semantics for g_ϕ. The variance weighting in Section 3.5 also consistently uses $\mathrm{Var}[\mathcal{L}_i]$ for each client without mixing notations.

Generalization Weight Recalibration

To further stabilize training, we introduce a *generalization adjustment mechanism* that reweights client contributions based on their generalization consistency. Each client is assigned a dynamic aggregation weight:

$$\alpha_i = \frac{\exp(-\tau\,\mathrm{Var}[\mathcal{L}_i])}{\sum_j \exp(-\tau\,\mathrm{Var}[\mathcal{L}_j])}, \tag{7}$$

where τ is a temperature parameter controlling the sharpness of weighting. Clients with lower loss variance (i.e., more stable generalization) receive higher aggregation weights, reducing the influence of noisy or unstable updates.

The server update at communication round t becomes:

$$\psi^{t+1} = \psi^t - \eta\sum_{i=1}^{N}\alpha_i\nabla_\psi\mathcal{L}_i^{(t)}, \tag{8}$$

where η is the learning rate.

This reweighting ensures that aggregation is guided by generalization reliability rather than raw performance.

Privacy Considerations: Although FedVR does not introduce a formal privacy guarantee, it reduces the exposure of raw client information by transmitting gradients of hypernetwork-generated parameters rather than task-specific model weights or features. This indirect communication layer makes gradient inversion more difficult in practice, as each client's update is mediated through a shared latent mapping (h_ψ, g_ϕ) instead of directly revealing instance-level gradients. Nevertheless, we acknowledge that full protection against gradient-based reconstruction attacks would require integrating formal differential privacy or secure aggregation mechanisms.

Experimental Setup

Datasets and evaluation protocol: We evaluate FedVR on three widely recognized benchmarks for federated domain generalization PACS (Li et al. 2017), Office-Home (Venkateswara et al. 2017), and VLCS (Fang, Xu, and Rockmore 2013) which collectively capture diverse distributional shifts and are thus well-suited for assessing cross-domain generalization under federated constraints. Following the standard DomainBed (Gulrajani and Lopez-Paz 2020) and FDG literature, we adopt a *leave-one-domain-out (LODO)* evaluation protocol: in each run, one domain is held out as the unseen *out-of-domain (OOD)* target, while the remaining domains are used for training. This process is repeated for all domains, and the average OOD accuracy is reported across targets, with each column in Table 1 corresponding to the accuracy when that domain serves as the unseen target. For implementation, we utilize a pretrained ResNet-50 backbone, where the hypernetwork generates adapter and classifier parameters while keeping the shared feature extractor frozen. This setup tests the scalability of our approach to deeper architectures and more expressive representations. Across all benchmarks, FedVR consistently enhances generalization to unseen domains, confirming its robustness to both model depth and capacity variations.

Training details and baselines: All methods are trained for 200 communication rounds, with each client performing two local epochs per round and using a batch size of 64. We report the mean and standard deviation over three random seeds for robustness. All hyperparameters, including the learning rate, variance regularization weight λ_{var}, and temperature τ, are tuned separately per dataset using the validation splits. We compare FedVR against representative federated and domain generalization baselines, including FedAvg (McMahan et al. 2017), FedProx (Li et al. 2020), and the hypernetwork-based personalization approach pFedHN (Shamsian et al. 2021). We also include two strong FDG methods: FedSR (Nguyen, Torr, and Lim 2022), which introduces feature-space regularization, and FedGMA (Tenison et al. 2022), which adaptively reweights clients. Additionally, we report results for two reference baselines: Centralized, where all source data are pooled for joint training, and Local-Only, where each client trains independently without collaboration. This comprehensive comparison ensures that improvements achieved by FedVR stem from its variance-regularized hypernetwork design rather than architectural or optimization biases.

Algorithm 1 provides a complete summary of FedVR, including client-side embedding computation, server-side hypernetwork updates, fairness-aware weighting, stability regularization, and zero-shot deployment.

Additional details and supplementary experiments: We report results using ResNet-50 in the main paper as the representative backbone. Results with other architectures (e.g., ResNet-18 and Vision Transformer), along with the full derivation of $\nabla_\psi L_{\text{var}}$, extended ablation studies, sensitivity analyses for λ_{var} and τ, and communication-efficiency evaluations, are provided in the Supplementary Material.

Table 1: Leave-one-domain-out (LODO) evaluation on PACS, OfficeHome, and VLCS. Each column reports the out-of-domain (OOD) accuracy when that domain is held out as the unseen target; μ denotes the mean OOD accuracy across all targets.

PACS					
Method	A	C	P	S	μ
Central	81.0	74.5	75.3	69.9	75.2
Local	75.7	70.0	69.7	62.1	69.4
FedAvg	82.4	75.2	78.5	70.8	76.7
FedProx	81.6	77.2	77.3	72.3	77.1
pFedHN	79.0	75.0	73.2	69.3	74.1
hFedF	81.4	74.9	75.6	71.5	75.9
FedVR (Ours)	**81.35**	**77.26**	**81.71**	**80.73**	**80.26**
OfficeHome					
Method	Art	Clip.	Prod.	Real	μ
Central	54.4	46.9	44.6	51.7	49.4
Local-Only	40.7	35.9	33.0	37.6	36.8
FedAvg	56.4	50.0	46.1	54.0	51.6
FedProx	56.4	48.8	46.2	55.6	51.8
pFedHN	49.1	44.5	46.5	46.2	46.6
hFedF	56.9	49.8	47.6	54.0	52.1
FedVR (Ours)	**64.28**	**53.40**	**65.49**	**67.71**	**62.72**
VLCS					
Method	C101	LMe	SUN	PAS	μ
Central	57.3	64.6	63.5	65.6	62.8
Local	54.9	58.9	61.6	59.6	58.8
FedAvg	61.4	65.1	66.7	67.5	65.2
FedProx	60.5	64.1	67.2	68.5	65.1
pFedHN	59.9	63.5	65.8	66.8	64.0
hFedF	62.2	66.2	66.4	67.7	65.6
FedVR (Ours)	**75.35**	**68.42**	**78.12**	**74.65**	**74.14**

These additional experiments and derivations confirm that FedVR maintains consistent improvements and theoretical clarity across model capacities and backbone types.

Results

Overall Performance: Table 1 summarizes the quantitative comparison of FedVR against state-of-the-art baselines across all benchmarks. Overall, FedVR consistently achieves the highest out-of-domain accuracy, reflecting its strong capacity to capture transferable knowledge under heterogeneous client distributions. Across PACS, OfficeHome, and VLCS, our approach demonstrates remarkable robustness to both stylistic and structural domain shifts. On the PACS benchmark characterized by large appearance variations among *Art*, *Cartoon*, *Photo*, and *Sketch* FedVR attains an average accuracy of 80.26%, surpassing hFedF by +4.36%. The substantial improvement on the most challenging *Sketch* domain highlights enhanced resilience to extreme style diversity. In the more complex OfficeHome dataset with 65 categories, FedVR achieves a mean accuracy of 62.72%, outperforming hFedF (52.1%) and FedProx by over +10.9%, with particularly strong results on the *Product* and *Real* domains where visual heterogeneity is greatest. Finally, on VLCS which blends scene and object-centric domains (*Caltech101*, *LabelMe*, *SUN09*, and *PASCAL VOC*) our model records 74.14% accuracy, exceeding the previous best baseline by +8.54%. These consistent margins across datasets validate the design of FedVR: variance-regularized aggregation mitigates overfitting to domain-specific outliers, while the hypernetwork-driven weight synthesis enables adaptive, domain-conditioned personalization that preserves global consistency and cross-domain fairness.

Conclusion

We introduced FedVR, a novel federated domain generalization framework that overcomes the convex-fusion ceiling of parameter-averaging methods through hypernetwork-driven parameter generation. Our approach learns a domain-conditioned hypernetwork that synthesizes client-specific models while enforcing cross-client consistency via variance regularization and generalization-weighted optimization. Extensive experiments on multiple datasets demonstrate state-of-the-art performance, with FedVR improving out-of-domain accuracy by up to 8.5%. Analysis confirms superior gradient alignment, tighter embedding distributions, and better calibration under domain shift.

Future work includes integrating formal differential privacy guarantees and extending our variance-harmonized framework to federated large-scale vision-language models. FedVR establishes a new paradigm for building generalizable models in decentralized, heterogeneous environments.

References

Bai, R.; Bagchi, S.; and Inouye, D. I. 2023. Benchmarking algorithms for federated domain generalization. *arXiv preprint arXiv:2307.04942*.

Bartholet, M.; Kim, T.; Beuret, A.; Yun, S.-Y.; and Buhmann, J. M. 2024. Hypernetwork-Driven Model Fusion for Federated Domain Generalization. *arXiv preprint arXiv:2402.06974*.

Carey, A. N.; Du, W.; and Wu, X. 2022. Robust personalized federated learning under demographic fairness heterogeneity. In *2022 IEEE International Conference on Big Data (Big Data)*, 1425–1434. IEEE.

Chen, X.; Huang, Y.; Xie, Z.; and Pang, J. 2024. HyperFedNet: Communication-efficient personalized federated learning via hypernetwork. *arXiv preprint arXiv:2402.18445*.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Fang, C.; Xu, Y.; and Rockmore, D. N. 2013. Unbiased metric learning: On the utilization of multiple datasets and web images for softening bias. In *Proceedings of the IEEE International Conference on Computer Vision*, 1657–1664.

Geiping, J.; Bauermeister, H.; Dröge, H.; and Moeller, M. 2020. Inverting gradients-how easy is it to break privacy in federated learning? *Advances in neural information processing systems*, 33: 16937–16947.

Gulrajani, I.; and Lopez-Paz, D. 2020. In search of lost domain generalization. *arXiv preprint arXiv:2007.01434*.

Guo, P.; Zeng, S.; Chen, W.; Zhang, X.; Ren, W.; Zhou, Y.; and Qu, L. 2025. A new federated learning framework against gradient inversion attacks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 16969–16977.

Gupta, S.; Sutar, V.; Singh, V.; and Sethi, A. 2025. FedAlign: Federated Domain Generalization with Cross-Client Feature Alignment. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 1801–1810.

Ha, D.; Dai, A.; and Le, Q. V. 2016. Hypernetworks. *arXiv preprint arXiv:1609.09106*.

Hatamizadeh, A.; Yin, H.; Molchanov, P.; Myronenko, A.; Li, W.; Dogra, P.; Feng, A.; Flores, M. G.; Kautz, J.; Xu, D.; et al. 2023. Do gradient inversion attacks make federated learning unsafe? *IEEE Transactions on Medical Imaging*, 42(7): 2044–2056.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Huang, Y.; Gupta, S.; Song, Z.; Li, K.; and Arora, S. 2021. Evaluating gradient inversion attacks and defenses in federated learning. *Advances in neural information processing systems*, 34: 7232–7241.

Karimireddy, S. P.; Kale, S.; Mohri, M.; Reddi, S.; Stich, S.; and Suresh, A. T. 2020. Scaffold: Stochastic controlled averaging for federated learning. In *International conference on machine learning*, 5132–5143. PMLR.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

Kumar, A.; et al. 2022. Bayesian Hypernetworks for Robust Multi-Domain Learning. In *Advances in Neural Information Processing Systems (NeurIPS)*.

Li, D.; Yang, Y.; Song, Y.-Z.; and Hospedales, T. 2018. Learning to generalize: Meta-learning for domain generalization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Li, D.; Yang, Y.; Song, Y.-Z.; and Hospedales, T. M. 2017. Deeper, broader and artier domain generalization. In *Proceedings of the IEEE International Conference on Computer Vision*, 5542–5550.

Li, H.; Cai, Z.; Wang, J.; Tang, J.; Ding, W.; Lin, C.-T.; and Shi, Y. 2023. FedTP: Federated learning by transformer personalization. *IEEE Transactions on Neural Networks and Learning Systems*, 35(10): 13426–13440.

Li, H.; Zhou, Z.; Cao, Z.; Li, X.; Chen, W.; and Zhang, X. 2025. FedSDAF: Leveraging Source Domain Awareness for Enhanced Federated Domain Generalization. *arXiv preprint arXiv:2505.02515*.

Li, T.; Sahu, A. K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; and Smith, V. 2020. Federated optimization in heterogeneous networks. *Proceedings of Machine Learning and Systems*, 2: 429–450.

Li, X.; Huang, K.; Yang, W.; Wang, S.; and Zhang, Z. 2019. On the convergence of fedavg on non-iid data. *arXiv preprint arXiv:1907.02189*.

Lin, Y.; Wang, H.; Li, W.; and Shen, J. 2023. Federated learning with hyper-network—A case study on whole slide image analysis. *Scientific Reports*, 13(1): 1724.

Liu, Q.; Chen, C.; Qin, J.; Dou, Q.; and Heng, P.-A. 2021a. FedDG: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1013–1023.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021b. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022.

Ma, S.; Xie, W.; Li, D.; Li, H.; and Li, Y. 2024. Reducing Spurious Correlation for Federated Domain Generalization. arXiv:2407.19174.

Ma, X.; Zhang, J.; Guo, S.; and Xu, W. 2022. Layer-wised model aggregation for personalized federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10092–10101.

McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; and y Arcas, B. A. 2017. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, 1273–1282. PMLR.

Nguyen, A. T.; Torr, P.; and Lim, S. N. 2022. FedSR: A simple and effective domain generalization method for federated learning. *Advances in Neural Information Processing Systems*, 35: 38831–38843.

Pourpanah, F.; Molahasani, M.; Soltany, M.; Greenspan, M.; and Etemad, A. 2025. Federated Unsupervised Domain Generalization using Global and Local Alignment of Gradients. arXiv:2405.16304.

Qu, J.; Faney, T.; Wang, Z.; Gallinari, P.; Yousef, S.; and de Hemptinne, J.-C. 2022. HMoE: Hypernetwork-based mixture of experts for domain generalization. *arXiv preprint arXiv:2211.08253*.

Radwan, A.; Soliman, M.; Abdelaziz, O.; and Shehata, M. 2025. FedDG-MoE: Test-Time Mixture-of-Experts Fusion for Federated Domain Generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1811–1820.

Rosenbaum, C.; Klinger, T.; and Riemer, M. 2017. Routing networks: Adaptive selection of non-linear functions for multi-task learning. *arXiv preprint arXiv:1711.01239*.

Seokeon, C.; Hyunsin, P.; Sungha, C.; et al. 2024. Client-agnostic learning and zero-shot adaptation for federated domain generalization. US Patent App. 18/238,998.

Shamsian, A.; Navon, A.; Fetaya, E.; and Chechik, G. 2021. Personalized federated learning using hypernetworks. In

International conference on machine learning, 9489–9502. PMLR.

Tashakori, A.; Zhang, W.; Wang, Z. J.; and Servati, P. 2023. SemiPFL: Personalized semi-supervised federated learning framework for edge intelligence. *IEEE Internet of Things Journal*, 10(10): 9161–9176.

Tenison, I.; Sreeramadas, S. A.; Mugunthan, V.; Oyallon, E.; Rish, I.; and Belilovsky, E. 2022. Gradient masked averaging for federated learning. *arXiv preprint arXiv:2201.11986*.

Venkateswara, H.; Eusebio, J.; Chakraborty, S.; and Panchanathan, S. 2017. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5018–5027.

Wang, J.; Liu, Q.; Liang, H.; Joshi, G.; and Poor, H. V. 2020. Tackling the objective inconsistency problem in heterogeneous federated optimization. *Advances in neural information processing systems*, 33: 7611–7623.

Yuan, J.; Ma, X.; Chen, D.; Wu, F.; Lin, L.; and Kuang, K. 2023. Collaborative semantic aggregation and calibration for federated domain generalization. *IEEE Transactions on Knowledge and Data Engineering*, 35(12): 12528–12541.

Zhang, D.; Chen, M.; Li, C.; Huang, L.; and Zhang, J. 2023a. Aggregation of Disentanglement: Reconsidering Domain Variations in Domain Generalization. arXiv:2302.02350.

Zhang, L.; Lei, X.; Shi, Y.; Huang, H.; and Chen, C. 2021. Federated learning with domain generalization. *arXiv preprint arXiv:2111.10487*.

Zhang, R.; Xu, Q.; Yao, J.; Zhang, Y.; Tian, Q.; and Wang, Y. 2023b. Federated domain generalization with generalization adjustment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3954–3963.

Zhou, Y.; Lv, J.; Tian, Y.; Si, D.; Ye, Q.; and Lv, J. 2025. Deploying Models to Non-participating Clients in Federated Learning without Fine-tuning: A Hypernetwork-based Approach. *arXiv preprint arXiv:2508.12673*.