FedTopo: Topology-Informed Representation Alignment in Federated Learning Under Non-I.I.D. Conditions

Ke Hu[1], **Liyao Xiang**[2], **Peng Tang**[1,3*], **Weidong Qiu**[1,3]
[1]School of Computer Science, Shanghai Jiao Tong University
[2]John Hopcroft Center for Computer Science, Shanghai Jiao Tong University
[3]Shanghai Key Laboratory of Integrated Administration Technologies for Information Security
crestiny@sjtu.edu.cn, xiangliyao08@sjtu.edu.cn, tangpeng@sjtu.edu.cn, qiuwd@sjtu.edu.cn

Abstract

Current federated-learning models deteriorate under heterogeneous (non-I.I.D.) client data, as their feature representations diverge and pixel- or patch-level objectives fail to capture the global topology which is essential for high-dimensional visual tasks. We propose **FedTopo**, a framework that integrates **Topological-Guided Block Screening (TGBS)** and **Topological Embedding (TE)** to leverage topological information, yielding coherently aligned cross-client representations by **Topological Alignment Loss (TAL)**. First, Topology-Guided Block Screening (TGBS) automatically selects the most topology-informative block, i.e., the one with maximal topological separability, whose persistence-based signatures best distinguish within- versus between-class pairs, ensuring that subsequent analysis focuses on topology-rich features. Next, this block yields a compact Topological Embedding, which quantifies the topological information for each client. Finally, a Topological Alignment Loss (TAL) guides clients to maintain topological consistency with the global model during optimization, reducing representation drift across rounds. Experiments on Fashion-MNIST, CIFAR-10, and CIFAR-100 under four non-I.I.D. partitions show that FedTopo accelerates convergence and improves accuracy over strong baselines.

Code — https://github.com/LonelyMoonDesert/MyGO-FL
Extended version — https://arxiv.org/abs/2511.12628

1 Introduction

Federated Learning (FL) enables decentralized training across clients. However, a central challenge is data heterogeneity—the *non-I.I.D.* problem. When local datasets are skewed, each client's gradient points toward a different optimum, and performance of the aggregated global model decreases. From a representational perspective, skewed local data is analogous to different teachers emphasizing distinct aspects of a concept. Each client learns a different interpretation of the data, leading to discrepancies in the learned feature representations. Recent studies (2023) have shown that this divergence is a key manifestation of non-I.I.D. data.

We illustrate this with a toy CIFAR-10 experiment. We select two clients: Client A holds cats and dogs, while Client B has disjoint cat samples and additional ships. Figure 1 illustrates feature inconsistency across clients from three perspectives: (1) **Feature distribution** (Figure 1(a)): We extract intermediate representations of "cat" samples and project them using t-SNE. The resulting clusters are client-specific with minimal overlap, indicating inconsistent representations even for the same class. (2) **Activation maps** (Figure 1(b)): For two inputs ("dog" and "cat"), we visualize the mean and max-channel activations. The maps reveal noticeable differences in spatial focus and intensity across clients. (3) **Topological structure** (Figure 1(c)): For a dog image (index 83), we compute persistence barcodes and diagrams. Discrepancies in H_0/H_1 features suggest mismatched topological summaries of the learned representations. This example shows that non-I.I.D. client datasets induce misalignment not only in parameter space but also in the deeper representational geometry and topology of features (2023). However, current methods generally ignore the latter.

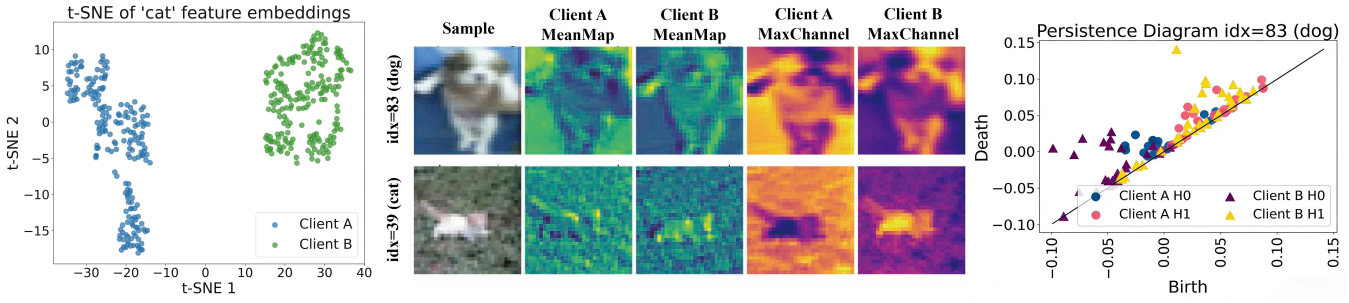
1.1 Contributions

We propose **FedTopo**, a topology-aware federated learning framework that enforces cross-client topological consistency to mitigate representation drift in non-I.I.D. scenarios. Our key contributions are:

- **Topology-Guided Block Screening (TGBS)**: A layer selection mechanism that identifies the block with maximal topological class separability, shown to correspond to the highest mutual information with class labels.
- **Topological Embedding (TE)**: A compact, persistence-based representation derived from the selected block, which is provably Lipschitz-stable under input and model perturbations.
- **Topological Alignment Loss (TAL)**: A regularizer that aligns each client's TE with that of the global model. We provide convergence analysis and introduce an adaptive weighting scheme based on a proposed non-I.I.D. metric and the temporal dynamics of topological loss.

FedTopo achieves both theoretical soundness and empirical effectiveness, consistently outperforming strong baselines on FMNIST, CIFAR-10, and CIFAR-100 under diverse non-I.I.D. conditions.

*Corresponding author.



(a) (b) (c)

Figure 1: Comparison of feature distributions, activation maps, and topological structure. (a) t-SNE of 'cat' feature embeddings. (b) Feature/activation map visualizations for two samples. (c) Persistence diagrams for the selected dog sample.

2 Related Work

2.1 Non-I.I.D. Federated Learning

Federated Learning (FL) enables collaborative training across decentralized clients while preserving data privacy (2025). However, performance degrades significantly under non-I.I.D. data due to representational drift and unstable convergence (2025). To address this, recent methods focus on improving the alignment between local and global models (2025). For instance, FedCCFA (2024) clusters client classifiers and aligns intermediate features to mitigate concept drift. FedSA (2025) introduces semantic anchors to regularize representation learning and alleviate feedback loops caused by statistical and model heterogeneity. A detailed summary of recent non-I.I.D. FL methods is provided in Appendix E.

2.2 Topological Data Analysis (TDA)

TDA provides global, multi-scale descriptors that are robust to noise, deformation, and transformation—properties well suited to the challenges of non-I.I.D. FL. Unlike pixel-level metrics (2024), topological invariants such as connected components and loops remain stable under continuous perturbations (2015). Prior work incorporates persistent homology into segmentation (2019; 2022), reconstruction, and classification tasks, yielding improved robustness and generalization (2021; 2024; 2021). These results suggest that topological structure can provide complementary supervision in heterogeneous learning environments.

2.3 Topological Alignment in Deep Learning

Topological alignment has been used to promote consistent representations across distributions and modalities. In graph learning, aligning higher-order structures—such as edge orbits—enhances embedding robustness (2022; 2023). Other work shows that neural networks can implicitly preserve topological invariants such as connectivity and loops (2025). In domain adaptation and change detection, explicitly aligning topological structure improves geometric consistency and classification boundaries (2021). These findings motivate our approach of aligning topological descriptors across clients to reduce representation drift in federated learning.

3 Preliminaries

FedTopo uses topological descriptors to summarize the topological information in feature maps across heterogeneous clients. This section introduces the topological tools for converting feature maps into compact topological descriptors, including cubical filtrations, persistence diagrams (PDs), and persistence images (PIs).

3.1 Feature Maps to Topological Descriptors

Figure 2 illustrates the full computation pipeline using a CIFAR-10 example.

Cubical Filtrations and Persistence Diagrams Given a local sample $x \in \mathcal{D}_i$, let $\phi_\ell(x; \mathbf{w}) \in \mathbb{R}^{C\times H\times W}$ denote the activation tensor at a selected layer ℓ of a local model $f(\cdot; \mathbf{w})$. We extract n_C channels, each yielding a 2D activation map $A \in \mathbb{R}^{H\times W}$ (Figure 2(a)).

To analyze the topology of A, we apply **cubical persistent homology** using a **sublevel set filtration**:

$$A^\lambda = \{(i,j) \in [H] \times [W] \mid A_{i,j} \leq \lambda\}. \tag{1}$$

As λ increases, more pixels are included, and topological features—such as components (H_0) and holes (H_1)—emerge and vanish. This produces a **persistence diagram** (Figure 2(b)) that records their birth and death thresholds (b_i, d_i).

From Persistence Diagrams to Persistence Images Persistence diagrams are expressive but difficult to use directly due to their non-Euclidean nature and variable size. To enable compatibility with learning systems, we adopt the **persistence image (PI)** (2017), a fixed-size, differentiable representation.

Each diagram is mapped from birth–death (b_i, d_i) to birth–persistence $(b_i, p_i = d_i - b_i)$ coordinates. A Gaussian kernel is then applied to each point to create a smooth density:

$$\rho(x,y) = \sum_i \exp\left(-\frac{(x-b_i)^2 + (y-p_i)^2}{2\sigma^2}\right), \tag{2}$$

which is rasterized onto a fixed grid and flattened into a vector $\psi_c(x; \mathbf{w}) \in \mathbb{R}^M$, the PI for channel c (Figure 2(c)). In Section 4.2, we aggregate these across channels to obtain the final *Topological Embedding*.

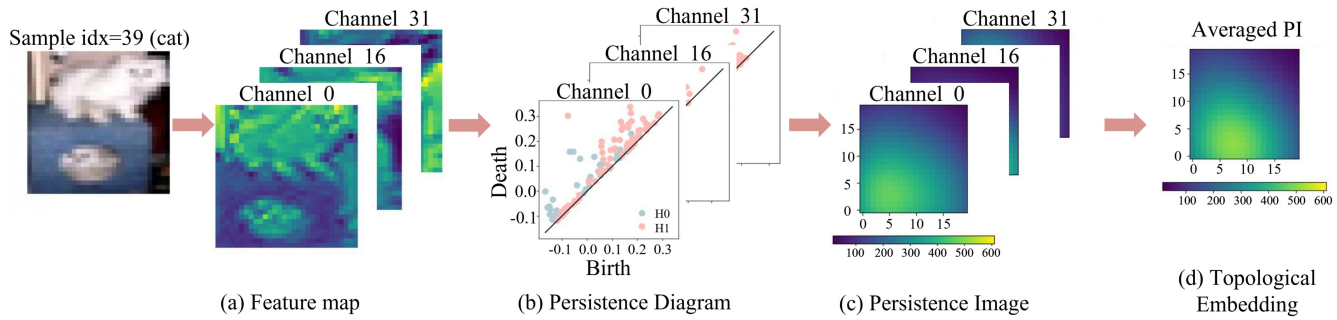


Figure 2: Topological embedding pipeline illustrated using a CIFAR-10 `cat` image. We show the input image, its intermediate feature map (a), the persistence diagrams from selected channels (b), the corresponding persistence images (c), and the final averaged topological embedding (d).

4 Proposed Method

We organize this section around the three components of FedTopo. First, we identify the most topology-informative block via Topology-Guided Block Screening (TGBS). Second, we extract a compact Topological Embedding (TE) from that block as a topological descriptor to represent each client's feature topology. Third, we introduce the Topological Alignment Loss (TAL) with an adaptive α scheduler to align local and global topologies and mitigate representation drift.

4.1 Topology-Guided Block Screening (TGBS)

Motivation We hypothesize that a block's topological signature, if it consistently distinguishes within-class and between-class samples, carries strong semantic signals. To target alignment where it matters most, FedTopo applies TGBS to identify the *most topology-informative block*—i.e., the one with highest topological class separability.

TGBS Design TGBS identifies the most topologically discriminative block by measuring its topological class separability (see Algorithm 1). For each candidate block, we extract activations, apply PCA for channel reduction, compute cubical persistence diagrams (H_0 and H_1), and convert them into fixed-size Persistence Images (PIs) for comparison (see Section 3.1 for details). We then calculate Euclidean distances (2017) between PIs and evaluate their ability to distinguish same-class vs. different-class sample pairs using ROC-AUC.

The block with the highest AUC is selected as topology-informative. In ResNet-18, TGBS selects `layer2` (AUC 0.814) as the most separable layer. Selections for other architectures are listed in Appendix B.

4.2 Topological Embedding (TE)

Given the per-channel persistence images ψ_c obtained in Section 3.1, we build a single descriptor for each sample by averaging over channels:

$$\mathbf{t}_i(x;\mathbf{w}) = \mathcal{T}_\ell(x;\mathbf{w}) = \frac{1}{K}\sum_{c=1}^{K}\psi_c(x;\mathbf{w}) \in \mathbb{R}^M. \qquad (3)$$

Algorithm 1: TGBS: Automatic Selection of the Topology-Informed Block

Input: Backbone f, candidate blocks $\mathcal{B}$, validation set $\mathcal{D}$, set of distance metrics $\mathcal{M}$, number of pairs N_{pairs}
Output: Topology-informative block $b^\star$ (and AUC score for each b)

for each block $b \in \mathcal{B}$ **do**
  Extract features $\{\mathbf{z}_i\} \leftarrow f_b(\mathcal{D})$
  Obtain embedding $\{\tilde{\mathbf{z}}_i\} \leftarrow \textsc{DimReduce}(\{\mathbf{z}_i\})$
  Compute persistence diagrams $\{\text{PD}_i\} \leftarrow \textsc{PersistentHomology}(\{\tilde{\mathbf{z}}_i\})$
  for each metric $m \in \mathcal{M}$ **do**
    Randomly sample N_{pairs} within-class pairs $\mathcal{P}_{\text{w}}$ and N_{pairs} between-class pairs $\mathcal{P}_{\text{b}}$
    for each $(j,k) \in \mathcal{P}_{\text{w}} \cup \mathcal{P}_{\text{b}}$ **do**
      Compute distance $d_{jk} \leftarrow \textsc{Dist}(\text{PD}_j, \text{PD}_k; m)$
      Set similarity $s_{jk} \leftarrow -d_{jk}$
    end for
    Calculate $\text{AUC}_m(b) \leftarrow \text{ROC\_AUC}(\{s_{jk}\})$
  end for
  Compute average $\text{AUC}(b) \leftarrow \frac{1}{|\mathcal{M}|}\sum_m \text{AUC}_m(b)$
end for
$b^\star \leftarrow \arg\max_{b\in\mathcal{B}} \text{AUC}(b)$
return $b^\star$ and $\{\text{AUC}(b)\}_{b\in\mathcal{B}}$

Here, K is the number of channels, and the resulting vector $\mathcal{T}_\ell$ is a compact topological descriptor of the activation at layer ℓ. Channel averaging removes permutation noise and reduces the dimension from $C \times H \times W$ to M, yielding a communication-efficient representation that underpins our Topological Alignment Loss (Section 4.3).

Computation cost. FedTopo introduces lightweight computation overhead. The most expensive step—computing topological embeddings via cubical persistent homology—has a worst-case time complexity of $O(C(HW)^\omega)$ per layer, with C channels, spatial size $H \times W$, and $\omega \leq 3$ (e.g., $\omega \approx 2.37$ for modern algorithms). In practice, efficient 2D cubical implementations with union-find and Morse heuristics yield near-linear runtime $\tilde{O}(CHW)$. For example, `layer2` in ResNet-18 ($C = 128$, $H = W = 8$)

involves $\sim 2.9 \times 10^4$ cells per block, incurring only a minor cost per batch as FedTopo requires just one additional forward of the selected block.

Communication cost. FedTopo does not add communication overhead during training, as all topological computations are local. If a single topological embedding of dimension M is uploaded per round (e.g., for monitoring), the overhead is

$$\Delta_{\text{FedTopo}} = \frac{M}{|\mathbf{w}|}. \tag{4}$$

With $M = 64$, this corresponds to 5.8×10^{-6} for ResNet-18 ($|\mathbf{w}| \approx 11$M) and 6.4×10^{-4} for a small CNN ($|\mathbf{w}| \approx 0.1$M), both well below 0.1% of model size. Thus, FedTopo preserves the communication efficiency of FedAvg.

4.3 Topological Alignment Loss (TAL) with Adaptive Weighting

The local objective for client i combines **cross-entropy** and **TAL**, which compares the client's descriptor to the global reference:

$$\mathcal{L}_i(\mathbf{w}_i; \bar{\mathbf{w}}) = \underbrace{\mathbb{E}_{(x,y)\sim\mathcal{D}_i}\left[\mathrm{CE}(f(x;\mathbf{w}_i), y)\right]}_{\text{cross-entropy}} + \alpha \underbrace{\mathbb{E}_{x\sim\mathcal{D}_i}\left[\left|\mathbf{t}_i(x;\mathbf{w}_i) - \mathbf{t}_i(x;\bar{\mathbf{w}})\right|^2\right]}_{\text{TAL}}, \tag{5}$$

where $\bar{\mathbf{w}}$ denotes the previous global model, and $\alpha > 0$ controls the regularization strength. The objective in Eq. (5) aligns local and global topological embeddings, reducing representational drift across clients.

Adaptive Scheduling of α. To modulate regularization dynamically, we set $\alpha_i = \lambda^{(r,e)} \cdot \alpha_i^{\text{base}}$, where $\lambda^{(r,e)}$ depends on global round and local epoch, and α_i^{base} reflects the current topology loss. We consider four scheduling strategies: (1) Warm-up, a static baseline without feedback; (2) Linear-Topo, which linearly scales α_i with the smoothed TAL loss; (3) Piecewise, which adjusts α_i using a piecewise-linear response to the loss; and (4) Smooth-Topo, which applies EWMA smoothing and a softened transformation for improved stability. See Appendix B.2 for details.

4.4 System Overview of FedTopo

FedTopo integrates three components into a unified topology-aware federated learning framework: (1) Topology-Guided Block Screening (TGBS) to identify the most topology-informative block, (2) construction of a compact Topological Embedding (TE) to describe the topology of feature maps, and (3) Topological Alignment Loss (TAL) with adaptive α scheduling to align local and global topological structures. Together, these components reduce representational drift and improve generalization under non-I.I.D. conditions.

Figure 3 illustrates the overall pipeline (Algorithm 2). In the pre-training stage (Figure 3(a)), TGBS selects the block with the highest class separability in topological space. During each communication round (Figure 3(b)), clients extract intermediate features from the selected block and compute the corresponding TE following the procedure described in Preliminaries. TAL is then applied alongside standard cross-entropy loss to update the local model. The adaptive α scheduler adjusts the strength of topological regularization across rounds to balance task performance and topological consistency.

Algorithm 2: FedTopo: Topology-Aware Federated Learning

Input: Initial global model $\bar{\mathbf{w}}^{(0)}$, client data $\{\mathcal{D}_i\}$, number of rounds R, TGBS-selected block $\ell^\star$
Output: Final global model $\bar{\mathbf{w}}^{(R)}$

Preprocessing (server-side):
  Run TGBS on validation data to select $\ell^\star$ (Algorithm 1)
for round $r = 1$ to R **do**
  Server broadcasts $\bar{\mathbf{w}}^{(r-1)}$ to all clients
  for each client i in parallel **do**
    Initialize local model $\mathbf{w}_i^{(r)} \leftarrow \bar{\mathbf{w}}^{(r-1)}$
    for local epoch $e = 1$ to E **do**
      Sample minibatch $(x, y) \sim \mathcal{D}_i$
      Compute cross-entropy loss $\mathcal{L}_{\text{CE}}$
      Compute $\mathbf{t}_i(x; \mathbf{w}_i)$ and $\mathbf{t}_i(x; \bar{\mathbf{w}}^{(r-1)})(Eq.\ (3))$
      Compute TAL $\mathcal{L}_{\text{TAL}}$ using Eq. (5)
      Update α_i using adaptive scheduler
      Update $\mathbf{w}_i$ via gradient step on total loss $\mathcal{L}_{\text{CE}} + \alpha_i \mathcal{L}_{\text{TAL}}$ using Eq. (5)
    end for
    Client sends updated $\mathbf{w}_i^{(r)}$ to server
  end for
  Server aggregates client models to update $\bar{\mathbf{w}}^{(r)}$
end for
return final model $\bar{\mathbf{w}}^{(R)}$

4.5 Theoretical Guarantees

We present key theoretical justifications for the design of FedTopo. Formal proofs of all propositions are provided in Appendix A.

AUC as a Proxy for Mutual Information TGBS selects the block with the highest between-class separability, measured by AUC. We formally justify this by showing that AUC is monotonic with mutual information:

Proposition 1 (AUC Implies Mutual Information)**.** *Let $T_b = T_b(X)$ be the topological embedding at block b, and define the distance distributions:*

$$d_b^+ \sim \left\{\|T_b(x) - T_b(x')\| \;\middle|\; Y(x) = Y(x')\right\},$$
$$d_b^- \sim \left\{\|T_b(x) - T_b(x')\| \;\middle|\; Y(x) \neq Y(x')\right\}. \tag{6}$$

Then AUC is defined as:

$$\mathrm{AUC}_b = \Pr_{u\sim d_b^+,\, v\sim d_b^-}\left[u < v\right]. \tag{7}$$

If $\mathrm{AUC}_{b_1} > \mathrm{AUC}_{b_2}$*, then* $I(T_{b_1}; Y) > I(T_{b_2}; Y)$*, implying*

$$\arg\max_b \mathrm{AUC}_b = \arg\max_b I(T_b; Y). \tag{8}$$

Lipschitz Stability of Topological Embedding TE is computed via channel-wise Persistence Images (PIs), which are known to be Lipschitz continuous (2017; 2017). This ensures the TE inherits input stability:

Proposition 2 (Stability of Topological Embedding)**.** *Let* $\mathbf{t}_i(x; \mathbf{w})$ *be the TE from* $\phi_\ell(x; \mathbf{w})$*. Then,*

$$\|\mathbf{t}_i(x; \mathbf{w}) - \mathbf{t}_i(x'; \mathbf{w})\|_2 \leq L \cdot \|x - x'\|_2. \tag{9}$$

This guarantees robustness of TAL to small input perturbations and supports stable convergence of the TAL loss.

Convergence Guarantee via Proximal Regularization We further show that TAL acts as a proximal regularizer, satisfying FedProx-like convergence properties (2020a; 2022):

Proposition 3 (FedProx-style Convergence under TAL)**.** *If* $F_i(w)$ *is L-smooth and the TAL term is* μ*-strongly convex, then the global objective converges to an* ε*-stationary point in* $O\left(\frac{L}{\mu}\log\frac{1}{\varepsilon}\right)$ *gradient steps.*

5 Experiments

We empirically evaluate FedTopo from multiple perspectives: overall performance across diverse non-I.I.D. settings, visualization of representation alignment, and ablation studies assessing the contributions of key components including TGBS, TAL, and adaptive α scheduling.

Data-partition schemes We consider four non-I.I.D. partition types: quantity skew, label skew, noise-based skew, and fixed-k label assignment.

Amplitude noise skew (N-skew). All clients share inputs x_i and add Gaussian noise $n_{i,j} \sim \mathcal{N}(0, \sigma_j^2 I)$. Let $\sigma_j = \frac{j-1}{n-1}\bar{\sigma}$ with $\sigma_n = 0$; then $\mathcal{D}_j = \{(x_i + n_{i,j}, y_i)\}$.

Label skew (L-skew). For each class c, normalize $p_{c,j}^{(\ell)}$ as $\hat{p}_{c,j} = \frac{p_{c,j}^{(\ell)}}{\sum_{m=1}^n p_{c,m}^{(\ell)}}$. The sample indices for client j are defined as $I_{c,j} = \left[\sum_{m<j}\hat{p}_{c,m}N_c, \sum_{m\leq j}\hat{p}_{c,m}N_c\right)$, and the dataset is $\mathcal{D}_j = \bigcup_c \{(x_i, y_i) \mid i \in I_{c,j}\}$.

Fixed-k label skew. Each client is assigned k labels from $[K]$ so that all labels are covered. For class c, the indices $\mathcal{I}_c$ are split into shards $\{\mathcal{I}_c^{(s)}\}$, and each client j is assigned data from classes in $\mathcal{C}_j$, yielding $\mathcal{D}_j = \bigcup_{c\in\mathcal{C}_j} \mathcal{I}_c^{(j,c)}$.

Quantity skew (Q-skew). For each client C_j, we sample proportions $\mathbf{p}^{(q)} \sim \mathrm{Dir}_n(\alpha_q)$ and normalize them as $\hat{p}_j = p_j^{(q)} / \sum_{m=1}^n p_m^{(q)}$. The sample indices for client j are then assigned by $I_j = \left[\sum_{m<j}\hat{p}_m N, \sum_{m\leq j}\hat{p}_m N\right)$, resulting in $\mathcal{D}_j = \{(x_i, y_i) \mid i \in I_j\}$.

Full definitions are provided in Appendix C.

Baselines We compare FedTopo with three families of non-I.I.D. FL methods: (1) **Aggregation regularization and momentum-based methods:** FedAvg (2017), FedAvgM (2020b), FedNova (2020), FedProx (2020b), SCAFFOLD (2020), and FedDyn (2023). These methods modify aggregation with regularization, correction terms, or momentum to mitigate client drift and accelerate convergence. (2) **Representation alignment and personalization methods:** MOON (2021), FedFTG (2022), FedPAC (2023), FedFM (2023), FedCCFA (2024), and FedSA (2025). They align representations and enable client-specific personalization. (3) **Server-side knowledge transfer methods:** FedDF (2020), FedDC (2022), and FedGen (2021). They use distillation or server-side synthetic data to reduce distribution shift.

Implementation Details We use mini-batch SGD (1951) for local optimization with $E = 5$ local epochs. All methods use a batch size of 32 and learning rate of 0.01. We train with $n = 5$ clients and run 10 communication rounds. For feature extraction, **FedTopo** uses `conv1` for the CNN on MNIST/FMNIST and `layer2` for ResNet-18 on CIFAR-10. Additional hyperparameters are listed in Appendix C.

5.1 Performance Evaluation

Accuracy and Convergence We run each method five times and report the average test accuracy of the global model in Table 1. FedTopo outperforms all baselines across all datasets and non-I.I.D. partition schemes. On FMNIST, CIFAR-10, and CIFAR-100, FedTopo achieves the best accuracy, improving over the second-best method by up to 8.69%, 11.05%, and 8.44%, respectively. These results demonstrate that FedTopo effectively addresses the challenges of non-I.I.D. data, reducing representation drift and improving feature alignment across clients.

Visualization To verify the benefits of aligning local models with the global feature space, we visualize the global model's representations using UMAP in Figure 4, with different colors for each client. Initially, in Round 0, the client features are distinct, with the global model far from the client distributions. By Round 2, client features begin to overlap with the global model, and by Round 9, the feature spaces nearly match, indicating successful alignment. We also visualize topological alignment using persistence barcodes in Appendix D, which show increasing similarity between client and global topologies. These results confirm that TAL promotes topological consistency and mitigates representation drift.

5.2 Ablation Study

We evaluate the contribution of each core component in **FedTopo** through three ablation studies, focusing on **Topology-Guided Block Screening (TGBS)**, **Topological Regularization via TAL** and **Adaptive Scheduling of** α.

w/o TGBS: Effect of Block Selection To assess the effectiveness of TGBS, we evaluate FedTopo using features extracted from different candidate blocks: (1) ResNet-18: {`conv1`, `layer1`, `layer2`, `layer3`, `layer4`}; (2)

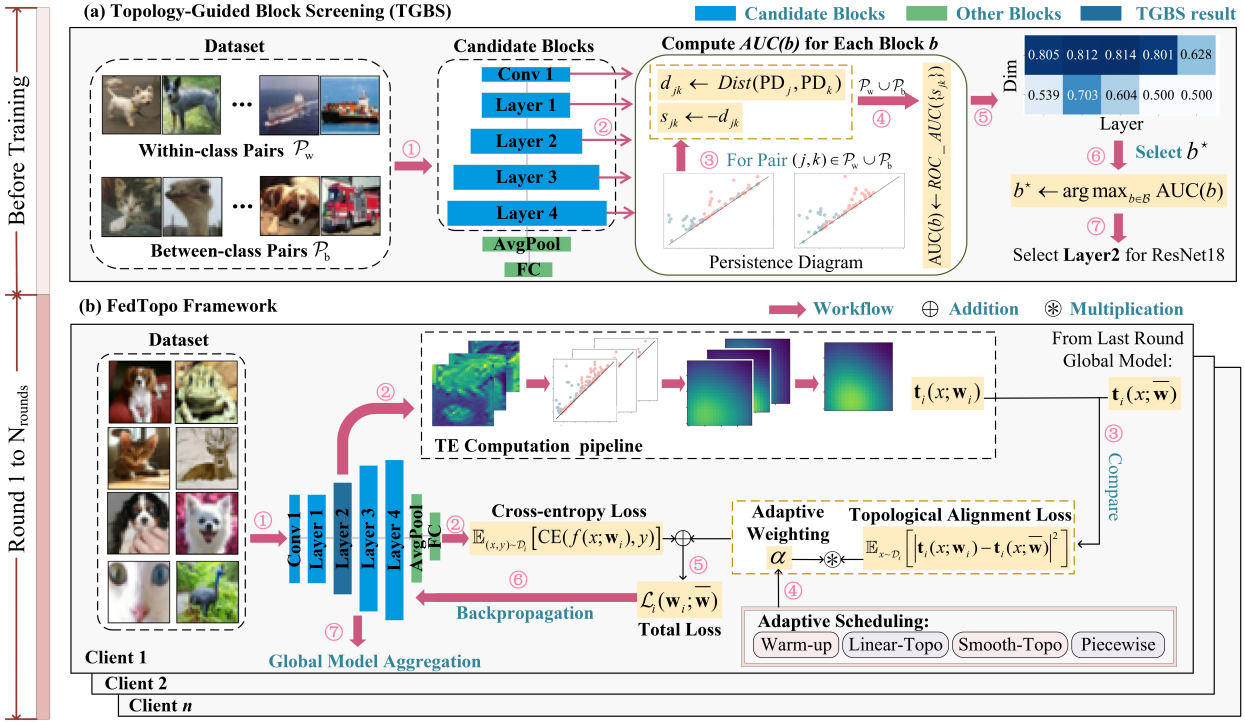


Figure 3: System overview of **FedTopo**, illustrated using ResNet-18 and CIFAR-10. (a) Before training, TGBS selects the most topology-informative block based on class separability in topological space. (b) During training, each client computes topological embeddings from intermediate features and applies Topological Alignment Loss (TAL) with adaptive scheduling.

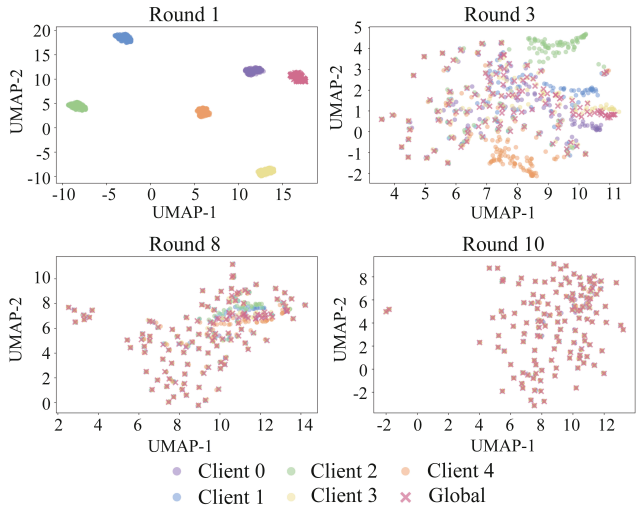


Figure 4: UMAP 2D feature projection of five clients and the global model at different communication rounds.

CNN: {`conv1`, `conv2`}. Table 2 shows the following: (1) In ResNet-18, `layer2` performs best, followed by `layer3`. (2) In CNN, `conv1` outperforms `conv2`, indicating that overly deep layers may degrade topological signals in shallow networks. (3) AUC aligns well with accuracy, validating topological separability as a criterion for block selection. TGBS thus selects `layer2` for ResNet-18 and `conv1` for CNN, both achieving the best results, confirming its effectiveness.

w/o TAL: Effect of Topological Regularization To evaluate the effect of Topological Alignment Loss (TAL), we compare FedTopo with and without TAL under three challenging non-I.I.D. scenarios: (A) ResNet-18 on CIFAR-10 with combined Q-skew ($\alpha_q = 0.5$) and K-skew ($k = 2$), (B) Simple CNN on FMNIST with severe L-skew ($\alpha_l = 0.1$), and (C) ResNet-18 on CIFAR-100 with combined L-skew ($\alpha_l = 0.5$) and N-skew ($\sigma = 1.0$). The global model accuracies are reported in Table 3. Across all settings, TAL yields substantial accuracy gains, averaging 13.25% (Setting A: +13.15%, Setting B: +14.80%, Setting C: +16.37%), showing that it reduces representation drift and enhances feature alignment in non-I.I.D. federated learning. These results demonstrate that TAL effectively reduces representation drift and enhances feature alignment in non-I.I.D. federated learning.

Effect of Adaptive α Scheduling Table 4 compares four α schedules—*Warm-up*, *Piecewise*, *Linear-Topo*, and *Smooth-Topo*—on two CIFAR-10 non-I.I.D. settings (A and B; see Section 5.2). We highlight three key observations. (1) In early rounds, *Linear-Topo* yields the fastest performance gain, improving Round-1 accuracy in Setting A by about +4% over Warm-up (69.15% vs. 65.13%), demonstrating

Dataset	FMNIST				CIFAR-10				CIFAR-100			
Non-I.I.D.	N	L	K	Q	N	L	K	Q	N	L	K	Q
FedAvg	60.01	83.83	75.97	91.07	65.10	86.47	66.69	90.62	46.87	60.53	49.35	63.43
FedAvg-M	65.43	87.70	80.21	92.86	67.87	89.76	66.85	93.39	48.87	62.83	49.47	65.37
FedNova	63.19	85.28	73.85	90.11	64.90	86.07	66.80	87.73	46.73	60.25	49.43	61.41
FedProx	66.68	87.49	75.91	90.57	63.47	85.85	66.99	89.35	45.70	60.10	49.57	62.55
SCAFFOLD	68.78	87.43	80.35	90.22	63.33	86.88	69.91	86.03	45.60	60.82	51.73	60.22
FedDyn	63.47	89.68	83.95	90.33	65.10	86.47	68.32	90.62	46.87	60.53	50.56	63.43
MOON	63.93	88.95	74.76	90.19	54.45	76.30	67.74	81.73	38.23	53.41	50.13	57.21
FedFTG	65.31	88.36	71.85	90.39	65.10	86.47	71.82	90.62	46.87	60.53	53.15	63.43
FedPAC	62.52	90.43	83.28	91.11	68.12	83.36	65.43	86.46	49.04	58.35	48.42	60.52
FedFM	64.68	86.28	81.94	90.02	64.68	87.43	69.44	91.09	46.57	61.20	51.39	63.76
FedCCFA	59.60	83.26	75.45	90.45	65.20	86.60	66.79	90.76	46.62	60.21	49.09	63.10
FedSA	71.47	90.40	85.22	92.25	67.59	88.08	71.75	91.58	48.67	61.95	53.10	64.11
FedDF	65.38	87.88	78.16	89.13	65.10	86.47	71.38	89.99	46.87	60.53	52.82	62.99
FedDC	68.78	91.08	86.46	90.78	66.99	87.03	70.71	89.22	48.23	60.92	52.33	62.46
FedGen	69.98	90.02	83.22	91.30	64.90	86.07	69.12	89.42	46.73	60.25	51.15	62.59
FedTopo	**74.96**	**91.30**	**89.88**	**94.46**	**73.87**	**92.76**	**77.89**	**96.63**	**53.20**	**65.93**	**57.64**	**67.64**

Table 1: Test accuracy (%) of global model under four types of non-IID partitioning. **N, L, K, Q** correspond to the following data partition types: (1) **N-skew (Amplitude Noise Skew)**: $\bar{\sigma} = 0.5$; (2) **L-skew (Label Skew)**: $\alpha_l = 0.1$; (3) **K-skew (Fixed-k Label Skew)**: $k = 5$; (4) **Q-skew (Quantity Skew)**: $\alpha_q = 0.5$.

Model	Blocks	AUC (%)	Acc. (%)
ResNet-18	conv1	80.54	80.62
	layer1	81.29	82.13
	layer2	**81.44**	**88.35**
	layer3	80.13	86.76
	layer4	62.80	85.35
CNN	**conv1**	**61.41**	**87.52**
	conv2	54.85	86.16

Table 2: Comparison of Topological Class Separability (AUC) and Federated Global Test Accuracy for Different Feature Blocks in ResNet-18 and CNN.

Methods	Setting A	Setting B	Setting C
w/o TAL	53.90	58.87	47.23
w/ TAL	**67.05**	**73.67**	**63.60**

Table 3: Test accuracy (%) of FedTopo with and without TAL under three representative non-I.I.D. settings.

Setting	Round	Warm	P-wise	Linear	Smooth
A	1	65.13	66.31	**69.15**	67.06
	10	73.52	74.72	75.27	**76.84**
B	1	58.32	59.48	**62.60**	60.86
	10	67.21	68.62	69.94	**71.31**

Table 4: Performance (%) of four α scheduling strategies on two CIFAR-10 non-I.I.D. partitions (Top-1 accuracy).

its effectiveness in quickly reducing cross-client representation gaps. (2) Over time, *Smooth-Topo* achieves the best final accuracy in both settings (76.84% / 71.31%) and shows the most stable training. (3) All topology-aware schedules outperform the naive Warm-up baseline, confirming the benefit of adaptive α scheduling.

6 Conclusion

This work introduces **FedTopo**, a topology-aware framework for federated learning under non-I.I.D. settings. By leveraging topological descriptors of intermediate feature maps, FedTopo aligns local representations with a global topological reference through **Topological Alignment Loss (TAL)**. To enable alignment, we propose **Topology-Guided Block Screening (TGBS)** for selecting the most discriminative layer, and construct a compact, differentiable **Topological Embedding (TE)** based on Persistence Images. Additionally, we incorporate an adaptive scheduling mechanism that dynamically adjusts regularization strength using topological feedback. Experiments on various datasets and partition schemes show that FedTopo effectively reduces representational drift and improves generalization. Theoretically, we establish the stability of TE and show that TAL preserves convergence under mild assumptions. Overall, our results highlight the potential of topological signals as a principled regularization for robust federated learning. Future work will extend FedTopo to graphs and sequences, develop privacy-preserving TE aggregation with online TGBS for large-scale deployments, and explore applications in decentralized data ecosystems, enabling secure data ownership verification, trading, and value extraction from multi-party data while preserving privacy.

Acknowledgments

This work was supported by the Shanghai Science and Technology Innovation Action Plan (24BC3200600) and National Natural Science Foundation of China (62441227).

References

Adams, H.; Chepushtanova, S.; Emerson, T.; Hanson, E.; Kirby, M.; Motta, F.; Neville, R.; Peterson, C.; Shipman, P.; and Ziegelmeier, L. 2015. Persistence Images: A Stable Vector Representation of Persistent Homology. *Journal of Machine Learning Research*, 16: 254–294.

Adams, H.; Emerson, T.; and Kirby, M. 2017. Persistence Images: A Stable Vector Representation of Persistent Homology. In *Journal of Machine Learning Research*, volume 18, 1–35.

Barannikov, S.; Trofimov, I.; Balabin, N.; and Burnaev, E. 2021. Representation topology divergence: A method for comparing neural network representations. *arXiv preprint arXiv:2201.00058*.

Carrière, M.; Cuturi, M.; and Oudot, S. 2017. Sliced Wasserstein Kernel for Persistence Diagrams. In *Proceedings of the 34th International Conference on Machine Learning*, 664–673.

Chen, J.; Li, Z.; Mao, H.; Woo, W. L.; and Peng, X. 2023. Cross-View Graph Consistency Learning for Invariant Graph Representations. *arXiv preprint arXiv:2311.11821*.

Clough, J.; Byrne, N.; Öksüz, I.; Zimmer, V.; Schnabel, J.; and King, A. 2022. A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44: 8766–8778.

Clough, J.; Öksüz, I.; Byrne, N.; Schnabel, J.; and King, A. 2019. Explicit topological priors for deep-learning based image segmentation using persistent homology. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 16–28. Springer.

Czolbe, S.; Feragen, A.; and Krause, O. 2021. Spot the Difference: Detection of Topological Changes via Geometric Alignment. In *NeurIPS 2021*.

et al., C. 2024. FedCCFA: Federated Learning with Classifier Clustering and Feature Alignment. In *NeurIPS 2024*.

et al., Z. 2025. FedSA: Federated Learning via Semantic Anchors. *arXiv*.

Gao, L.; Fu, H.; Li, L.; Chen, Y.; Xu, M.; and Xu, C.-Z. 2022. Feddc: Federated learning with non-iid data via local drift decoupling and correction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10112–10121.

Halverson, J.; and Ruehle, F. 2025. Learning Topological Invariance. *Preprint*.

Hu, e. 2021. Topology-aware Segmentation Using Discrete Morse Theory. In *International Conference on Learning Representations (ICLR)*. Persistent-homology inspired topological loss for segmentation.

iQua+. 2025. Federated Learning: Data Privacy and Non-IID Challenges. *arXiv*.

Jin, C.; Chen, X.; Gu, Y.; and Li, Q. 2023. FedDyn: A dynamic and efficient federated distillation approach on Recommender System. In *2022 IEEE 28th International Conference on Parallel and Distributed Systems (ICPADS)*, 786–793. IEEE.

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.

Li, Q.; He, B.; and Song, D. 2021. Model-contrastive federated learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 10713–10722.

Li, T.; Sahu, A. K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; and Smith, V. 2020a. Federated Optimization in Heterogeneous Networks. *Proceedings of the 36th International Conference on Machine Learning (ICML)*.

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

Lin, T.; Kong, L.; Stich, S. U.; and Jaggi, M. 2020. Ensemble distillation for robust model fusion in federated learning. volume 33, 2351–2363.

Ma, M.; Li, T.; and Peng, X. 2024. Beyond the federation: Topology-aware federated learning for generalization to unseen clients. *arXiv preprint arXiv:2407.04949*.

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.

MDPI+. 2025. Federated Learning: Approaches and Challenges in Mitigating Data Heterogeneity. *arXiv*.

Robbins, H.; and Monro, S. 1951. A stochastic approximation method. *The annals of mathematical statistics*, 400–407.

Sun, Q.-X.; Lin, X.; Zhang, Y.; Zhang, W.; and Chen, C. 2022. Towards Higher-order Topological Consistency for Unsupervised Network Alignment. *arXiv preprint arXiv:2208.12463*.

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.

Xu, J.; Tong, X.; and Huang, S.-L. 2023. Personalized federated learning with feature alignment and classifier collaboration. *arXiv preprint arXiv:2306.11867*.

Ye, R.; Ni, Z.; Xu, C.; Wang, J.; Chen, S.; and Eldar, Y. C. 2023. FedFM: Anchor-Based Feature Matching for Data Heterogeneity in Federated Learning. *IEEE Transactions on Signal Processing*, 71: 4224–4239.

Yuan, X.-T.; and Li, P. 2022. On Convergence of FedProx: Local Dissimilarity Invariant Bounds, Non-smoothness and Beyond. *arXiv preprint arXiv:2206.05187*.

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

the IEEE/CVF conference on computer vision and pattern recognition, 10174–10183.

Zhu, Z.; Hong, J.; and Zhou, J. 2021. Data-free knowledge distillation for heterogeneous federated learning. 12878–12889.

Zia, A.; Khamis, A.; Nichols, J.; Tayab, U. B.; Hayder, Z.; Rolland, V.; Stone, E.; and Petersson, L. 2024. Topological Deep Learning: a Review of an Emerging Paradigm. *Artificial Intelligence Review*, 57: 77.