Enhanced Federated Deep Multi-View Clustering Under Uncertainty Scenario

Bingjun Wei[1,2*], **Xuemei Cao**[1,2*], **Jiafen Liu**[1,2†], **Haoyang Liang**[1,2], **Xin Yang**[1,2†]

[1]School of Computing and Artificial Intelligence, Southwestern University of Finance and Economics
[2]Cognitive Computing and Crowd Intelligence Lab, Southwestern University of Finance and Economics
223081200006@smail.swufe.edu.cn, caoxuemei.qpz@gmail.com, jfliu@swufe.edu.cn, 223081200042@smail.swufe.edu.cn, yangxin@swufe.edu.cn

Abstract

Traditional Federated Multi-View Clustering assumes uniform views across clients, yet practical deployments reveal heterogeneous view completeness with prevalent incomplete, redundant, or corrupted data. While recent approaches model view heterogeneity, they neglect semantic conflicts from dynamic view combinations, failing to address dual uncertainties: view uncertainty (semantic inconsistency from arbitrary view pairings) and aggregation uncertainty (divergent client updates with imbalanced contributions). To address these, we propose a novel Enhanced Federated Deep Multi-View Clustering framework: first align local semantics, hierarchical contrastive fusion within clients resolves view uncertainty by eliminating semantic conflicts; a view adaptive drift module mitigates aggregation uncertainty through global-local prototype contrast that dynamically corrects parameter deviations; and a balanced aggregation mechanism coordinates client updates. Experimental results demonstrate that EFDMVC achieves superior robustness against heterogeneous uncertain views across multiple benchmark datasets, consistently outperforming all state-of-the-art baselines in comprehensive evaluations.

Introduction

Multi-view Clustering (MVC) (Fang et al. 2023; Liang et al. 2026) improves clustering accuracy by fusing complementary and diverse information from heterogeneous views (e.g., images, texts, and videos). However, its centralized implementations require aggregating raw data into a central server, posing risks of data breaches (Li et al. 2022) and violating privacy regulations (Cao et al. 2022). Federated Learning (FL) (Fan et al. 2025; Yang et al. 2024), as a distributed collaborative framework, offers a new paradigm for decentralized processing of multi-view data, achieving privacy preservation through local model training and global parameter aggregation.

The combination of FL and MVC, termed Federated Multi-View Clustering (FedMVC) (Huang et al. 2022; Yang and Sinaga 2025), aims to explore more comprehensive clustering structures from unsupervised multi-view data distributed across multiple clients while preserving privacy.

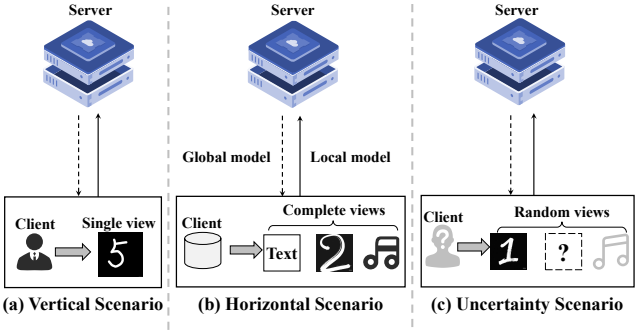

Figure 1: Three scenarios of FedMVC

Current FedMVC methods seek to learn a global model from multi-view data distributed across multiple devices, integrating deep learning (Huang et al. 2022) and matrix factorization (Wu et al. 2024) to mine complementary information from views across clients. Through FL aggregation mechanisms (McMahan et al. 2017), these methods consolidate models on the server side to derive a robust global model.

As shown in Fig. 1, existing FedMVC methods generally presuppose a fixed view distribution across clients: either every client holds a single view (Jiang et al. 2024) or all clients possess the complete set of views (Che et al. 2022). While such assumptions are workable under ideal conditions, they falter in real-world uncertain environments. FMCSC (Chen et al. 2024) does explore heterogeneous hybrid views, yet it is confined to the coexistence of the two static configurations above and cannot adapt to scenarios where the number of views per client changes dynamically, which severely limits its generalizability. In practice, clients may hold any number and combination of views: Hospital A has CT, MRI, and textual records simultaneously; Hospital B can only provide CT scans; Hospital C temporarily lacks MRI due to equipment maintenance and thus retains only CT and textual data. Such uncertain view distributions introduce multiple challenges:

View Uncertainty: Clients hold varying view combinations. multi-view clients face feature conflicts from heterogeneous views; single-view clients suffer representation bias, degrading global clustering convergence.

Aggregation Uncertainty: Inconsistent view structures across clients undermine the stability of federated optimiza-

*These authors contributed equally.
†Co-Corresponding author.

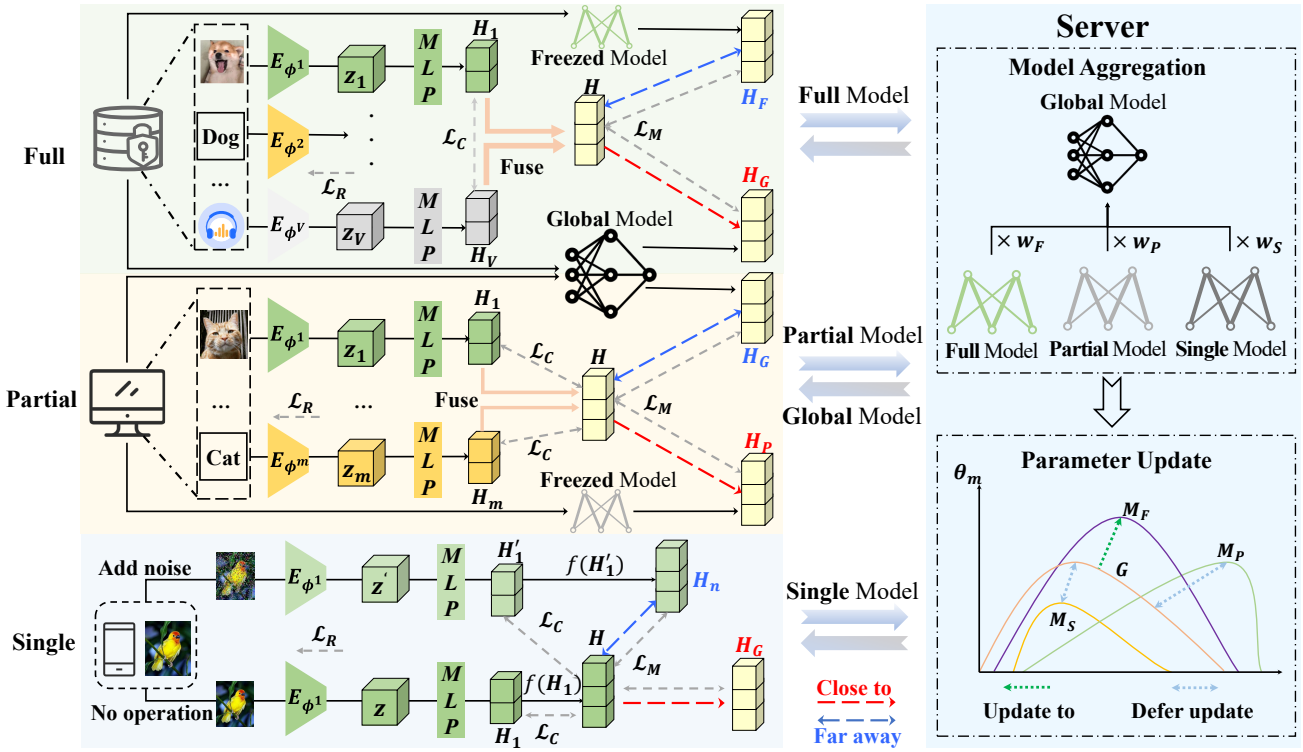

Figure 2: In the EFDMVC approach, for each client, $\mathbf{z}$ represents the learned feature representation, $\mathbf{H}$ denotes the local semantic features of the client, $\mathbf{H}_G$ stands for the global model features extracted based on client data, and the **Freezed Model** is preserved from the local model state saved at the end of the previous training round.

tion, making global model aggregation less robust.

To tackle the challenges outlined above, we propose Enhanced Federated Deep Multi-View Clustering (EFDMVC), as illustrated in Fig. 2. EFDMVC is tailored for uncertain federated environments, where clients hold arbitrary subsets of views and face highly heterogeneous data distributions. We begin with feature initialization to filter out irrelevant information and align local semantic spaces. It then employs a hierarchical contrastive fusion strategy to alleviate the impact of view-specific noise and capture shared semantics across clients. Especially for single-view clients, we generate synthetic noisy samples to enable contrastive learning. To further enhance model consistency, we introduce a view adaptive drift mechanism that recalibrates update directions and mitigates model divergence. Additionally, we develop a view balanced aggregation strategy that dynamically adjusts client weights and harmonizes parameter update magnitudes across diverse view configurations, effectively reducing aggregation uncertainty. Our primary contributions are summarized as follows:

- We propose the EFDMVC method, which is the first time to model the challenge of the client dynamically holding arbitrary view subsets and highly heterogeneous data in the real scene.
- We eliminated feature conflicts through hierarchical contrastive learning. And further propose an adaptive drift compensation module and adopt balanced aggregation to consider view quality and dynamically optimize the aggregation process.
- Experimental analyses validate EFDMVC effectiveness, demonstrating superior clustering performance across various uncertainty scenarios.

Related Work

Federated Multi-View Clustering

As shown in Fig. 1, existing FedMVC methods are generally categorized based on data partitioning:

(1) Traditional FedMVC relies on conventional machine learning, employing static feature extraction and rigid aggregation rules. Fed-MVKM(**?**) alternately optimizes multi-view centroids between clients and server, while SFOMVC-TR(Feng et al. 2025) innovatively applies Tucker decomposition to enable mixed client participation with factor matrix transmission.

(2) Vertical FedMVC assumes that V views are distributed to V single-view clients with shared samples. FedDMVC (Chen et al. 2023) employs a global self-supervised contrastive framework for privacy-preserving view alignment, while HFMVC (Jiang et al. 2024) enhances semantic consistency via a heterogeneity-aware dual contrastive mechanism.

(3) Horizontal FedMVC assumes multiple multi-view clients with non-overlapping samples. FMCSC (Chen et al. 2024) supports heterogeneous architectures with full-view

and single-view clients by maximizing cross-client mutual information.

Despite their effectiveness, these methods struggle with more realistic scenarios involving uncertainty hybrid views, where the number and quality of views vary across clients. EFDMVC addresses this by bridging both view and aggregation uncertainty under a hybrid horizontal–vertical setting.

Contrastive Learning

Contrastive Learning (CL) learns discriminative representations by constructing positive and negative pairs in unsupervised settings. Representative methods include SimCLR(Chen et al. 2020), which uses data augmentation and contrastive loss to structure the feature space; MoCo(He et al. 2020), which introduces a momentum encoder and a negative queue to alleviate sample scarcity; and BYOL(Grill et al. 2020), which eliminates explicit negatives by aligning online and target network outputs.

Integrating CL into Multi-View Clustering (MVC) and Federated Learning (FL) poses challenges in data heterogeneity and privacy. In MVC, methods like MFLVC(Xu et al. 2022) leverage cross-view and label-level contrast to extract shared semantics while suppressing view-specific noise. In FL, MOON addresses model drift from *Non-IID* data via model-level contrast, while FedX(Li, He, and Song 2021) employs cross-client knowledge distillation to jointly handle heterogeneity and privacy concerns.

Method

Problem Setting

In a uncertainty FedMVC environment, the C clients are categorized into three distinct types: single-view, partial-view, and full-view participants. Each client can have a maximum of V views, and the number of non-overlapping samples per client can be approximated as N. The clients provide raw feature data, where multiple views share K common clustering patterns to be discovered.

Each single-view client $\mathbf{s}$ possesses a dataset $\mathcal{D}_s = \{\mathbf{x}_i^{v_s}\}_{i=1}^N$, where $v_s \in \{1, \ldots, V\}$ denotes the randomly assigned view type for that client. For partial-view client $\mathbf{p}$, have dataset $\mathcal{D}_p = \{\mathbf{x}_i^{v \in \mathcal{V}_p}\}_{i=1}^N$ containing samples from a subset of views $\mathcal{V}_p \subset \{1, \ldots, V\}$ (where $|\mathcal{V}_p| \geq 2$). Full-view client $\mathbf{f}$ maintain complete multi-view dataset $\mathcal{D}_f = \{(x_i^1, \ldots, x_i^V)\}_{i=1}^N$. The client model $M_C(\cdot)$ maps heterogeneous inputs to a unified space $\mathbb{R}^h$. After $\mathbf{r}$ communication rounds, the server aggregates local models to produce global model $G(\cdot)$.

Feature Alignment Initialization

To address feature redundancy in heterogeneous multi-view data and cross-client view inconsistencies, we align semantic spaces across clients using autoencoders:

Each client processes its raw view data through view-specific encoders (Hinton, Krizhevsky, and Wang 2011), to obtain latent features $z_i^v = E_{\phi^v}(x_i^v) \in \mathbb{R}^{d_v}$:

$$\mathcal{L}_R = \frac{1}{N} \sum_{v \in \mathcal{V}'} \sum_{i=1}^{N} \|x_i^v - D_{\theta^v}(z_i^v)\|_2^2, \tag{1}$$

where $\mathcal{D}$ represents the local dataset ($\mathcal{D}_f$, $\mathcal{D}_p$, or $\mathcal{D}_s$), and $\mathcal{V}'$ denotes the set of available views for the client.

For each client, the features $\{\mathbf{z}^v\}_{v=1}^{|\mathcal{V}'|}$, obtained via Eq. (1), contain a mixture of common semantics and view-private information. To extract higher-level representations, we treat $\{\mathbf{z}^v\}_{v=1}^{|\mathcal{V}'|}$ as low-level features and further process them using a feature **MLP**. This yields the high-level features $\{\mathbf{h}^v\}_{v=1}^{|\mathcal{V}'|}$, where each $\mathbf{h}^v \in \mathbb{R}^h$, we construct a nonlinear mapping $\mathcal{H}(\mathbf{Z}; \Psi) : \mathbb{R}^{\sum_{v=1}^{|\mathcal{V}'|} d_v} \to \mathbb{R}^h$ defined as:

$$\mathbf{H} = \mathcal{H}(\mathbf{Z}; \Psi) = \mathcal{H}\left(\mathbf{Z}^1, \mathbf{Z}^2, \ldots, \mathbf{Z}^{|\mathcal{V}'|}; \Psi\right), \tag{2}$$

where $\{\mathbf{h}_i\}_{i=1}^N = \mathbf{H} \in \mathbb{R}^{N \times d}$, and $\mathbf{Z} \in \mathbb{R}^{N \times \sum_{v=1}^{|\mathcal{V}'|} d_v}$. Our objective is to preserve the discriminative power of the low-level features to avoid model collapse while learning the shared semantic representation $\mathbf{H}$ across views in the high-level feature space.

Hierarchical Contrastive Fusion

We propose a hierarchical contrastive learning fusion to address the issue of insufficient semantic consistency between different clients and adapt to heterogeneous view configurations. This contrastive framework enables each client to adverse effects of view-private information and learn consistent semantics while preserving local privacy. Specifically, it operates at three granularity levels:

Full-view client: It represents $\mathcal{V}' = \{1, \ldots, V\}$ and $\mathcal{D} = \mathcal{D}_f$. Each high-level feature $\mathbf{h}_i$ has $(NV - 1)$ feature pairs, i.e. , $\{\mathbf{h}_i^v, \mathbf{h}_j^n\}_{\substack{n=1,\ldots,V \\ j=1,\ldots,N}}$, where $\{\mathbf{h}_i^v, \mathbf{h}_i^n\}_{v \neq n}$ are $(N-1)$ positive feature pairs and the remaining $N(V-1)$ feature pairs are negative feature pairs.

The similarities of positive pairs should be maximized, while those of negative pairs should be minimized. Inspired by NT-Xent (Chen et al. 2020), the cosine distance is applied to measure the similarity between two features:

$$d(\mathbf{a}, \mathbf{b}) = \frac{\langle \mathbf{a}, \mathbf{b} \rangle}{\|\mathbf{a}\| \|\mathbf{b}\|}, \tag{3}$$

where $\langle \cdot, \cdot \rangle$ is dot product operator. Then, the feature contrastive loss between $\mathbf{h}_i^v$ and $\mathbf{h}_i^n$ is formulated as:

$$\mathcal{L}_C^f = -\frac{1}{N} \sum_{v=1}^{V} \sum_{i=1}^{N} \log \frac{e^{d(\mathbf{h}_i^v, \mathbf{h}_i^n)/\tau}}{\sum_{j=1}^{N} \sum_{v'=v,n} e^{d(\mathbf{h}_i^v, \mathbf{h}_j^{v'})/\tau} - e^{1/\tau}}, \tag{4}$$

where τ denotes the temperature parameter.

Similar to learning the high-level features, we adopt CL to achieve this consistency objective. For the v-th view, the same cluster labels $\mathbf{Q}_j^v = softmax(\mathbf{h}_j^v)$ have $(VK-1)$ label pairs, i.e., $\{\mathbf{Q}_j^v, \mathbf{Q}_k^n\}_{\substack{n=1,\ldots,V \\ k=1,\ldots,K}}$, where $\{\mathbf{Q}_j^v, \mathbf{Q}_j^n\}_{n \neq v}$ are constructed as $(V-1)$ positive label pairs, and the remaining $V(K-1)$ label pairs are negative label pairs.

We further define the label contrastive loss between $\mathbf{Q}_i^v$

and $\mathbf{Q}_i^n$ as:

$$\mathcal{L}_C^l = \frac{1}{K}\sum_{v=1}^{V}\sum_{j=1}^{K}\log\frac{-e^{d(\mathbf{Q}_j^v,\mathbf{Q}_j^n)/\tau}}{\sum_{k=1}^{K}\sum_{v'=v,n}e^{d(\mathbf{Q}_j^v,\mathbf{Q}_k^{v'})/\tau}-e^{1/\tau}} + \sum_{v=1}^{V}\sum_{j=1}^{K}s_j^v\log s_j^v, \tag{5}$$

where $s_j^v = \frac{1}{N}\sum_{i=1}^{N} q_{ij}^v$, q_{ij}^v represents the probability that the i-th sample belongs to the j-th cluster in the v-th view.

Partial-view client: $\mathcal{V}' = \mathcal{V}_p$ and $\mathcal{D} = \mathcal{D}_p$. It only contain a subset of the complete views, directly performing cross-view comparison faces the challenge of insufficient information. To address this, we propose a local contrastive learning module that aligns partial-view features $\mathbf{h}_i^v$ with common semantics $\mathbf{H}_i^p$. By establishing local-global correspondences, partial-view clients can leverage global common semantics to compensate for their perspective limitations, thereby obtaining more discriminative feature representations, the contrastive loss between $\mathbf{h}_i^v$ and $\mathbf{H}_i^p$ defined as:

$$\mathcal{L}_C^p = -\frac{1}{N}\sum_{v=1}^{|\mathcal{V}_p|}\sum_{i=1}^{N}\log\frac{e^{d(\mathbf{H}_i^p,\mathbf{h}_i^v)/\tau}}{\sum_{j\neq i}e^{d\left(\mathbf{H}_i^p,\mathbf{h}_j^v\right)/\tau}}, \tag{6}$$

where $\mathbf{h}_i^p$ is fused by the representations $\{\mathbf{h}_i^v\}_{v=1}^{|\mathcal{V}_p|}$.

Single-view client: $\mathcal{V}' = \{v_p\}$ and $\mathcal{D} = \mathcal{D}_s$. The absence of multi-view data necessitates an alternative contrastive module. We propose a noise-enhanced paradigm that establishes virtual negative pairs (Chen et al. 2020) through perturbed samples while aligning local features with global prototypes. Specifically, we generate perturbed samples $\mathbf{h}_i^{'}$ by adding noise to local features $\mathbf{h}_i$, then contrast them with common semantics $\mathbf{H}_i^s$ using the following loss function:

$$\mathcal{L}_C^s = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{d(\mathbf{H}_i^s,\mathbf{h}_i)/\tau}}{e^{d(\mathbf{H}_i^s,\mathbf{h}_i)/\tau}+e^{d(\mathbf{h}_i,\mathbf{h}_i^{'})/\tau}}, \tag{7}$$

where $\mathbf{h}_i^s$ is obtained via a nonlinear transformation of the single-view representation $\mathbf{h}_i$, and $\mathbf{h}_i^{'}$ represents locally perturbed features generated through additive noise $\mathbf{h}_i^{'} = M_s(x_i + \epsilon_i)$ with $\epsilon_i \sim \mathcal{N}(0, \sigma^2\mathbf{I})$.

View Adaptive Drift

Inspired by MOON (Li, He, and Song 2021; Huang et al. 2023), the model $M_f(\cdot)$ trained on the full-view client possesses the most data, while the globally aggregated model $G(\cdot)$ incorporates knowledge from all clients. we lveraging the global model broadcast at every communication round, and devise a view adaptive drift scheme: by carefully crafting diverse positive and negative sample pairs, we enable fine-grained exchange of client-specific knowledge, markedly suppressing semantic conflicts arising from missing views, while a regularization term steers the update direction to precisely mitigate aggregation uncertainty.

Therefore, to address the issue of view uncetainty among models, we employ CL to achieve specific model drifts for the three types of clients. The formulation is as follows:

$$\mathcal{L}_M = -\log\frac{e^{d(\mathbf{H},\mathbf{H}^+/\tau)}}{e^{d(\mathbf{H},\mathbf{H}^+)/\tau}+e^{d(\mathbf{H},\mathbf{H}^-)/\tau}} + \frac{\mu}{2}||\omega_M - \omega_G||^2, \tag{8}$$

where μ controls the strength of the proximal regularization term, $\mathbf{H}^+$ represents the positive sample, and $\mathbf{H}^-$ represents the negative sample. ω_M and ω_G are the parameters of the local model and global model , respectively.

In the r-th round of training, for the full-view client **f**, which is trained on complete multi-view data, its parameter update direction should be aligned with its own local model. Here, $\mathbf{H}^+$ is defined as the common semantics $\mathbf{H}^f = M_f^{(r-1)}(\mathcal{D}_f)$, while $\mathbf{H}^-$ is given by the global model feature $\mathbf{H}^g = G^{(r-1)}(\mathcal{D}_f)$. For partial-view client **p** and single-view client **s**, their local models are susceptible to data bias. The best available model they can access is $G^{(r-1)}(\cdot)$. Therefore, for these clients, $\mathbf{H}^+$ is set to $\mathbf{H}^g = G^{(r-1)}(\mathcal{D})$, while H^- is set to their respective local models from the previous round, i.e., $\mathbf{H}^p = M_p^{(r-1)}(\mathcal{D}_p)$ for client **p** and the noise fusion feature $\mathbf{H}' = M_s^{(r)}(\mathcal{D}_s')$ for client **s**, where $\mathcal{D}_s'$ represents the noise dataset. This module effectively addresses model drift for all three types of clients.

Balanced View Aggregation

Faced with the vast uncertainty of arbitrary view combinations in uncertain scenarios, conventional federated aggregation paradigms fall short; we must instead design an aggregation mechanism that simultaneously accounts for view complementarity and data trustworthiness, thereby achieving a stable and controllable global aggregation.

After completing local model training, each client uploads its trained model parameters to the central server. On the server side, a balanced view aggregation strategy is employed, which takes into account both the training sample size and multi-view data completeness of each participating node. This uncertainty weighted fusion mechanism ensures that the global model converges toward the optimal direction:

$$w_k = \frac{\alpha_c \cdot n_c}{\sum_{i=1}^{C}(\alpha_i \cdot n_i)}, \tag{9}$$

where n_c denotes the sample size of client c , and α_c prioritizes clients with richer view coverage. This design ensures that clients possessing both sufficient data and complete views exert greater influence during aggregation.

The global model ω_G is then updated by harmonizing all local models:

$$\omega_G^{(r)} = \sum_{c=1}^{C} w_c \cdot \omega_c^{(r)}, \tag{10}$$

where $\omega_c^{(r)}$ represents the parameters of client c in the communication round r. By adaptively balancing data volume and view quality, this mechanism mitigates biases caused by skewed view distributions and drives the global model toward a Pareto-optimal state.

Algorithm 1: Pipeline of EFDMVC

Require: Dataset with N samples and V views distributed among C clients, with an expectation to be partitioned into K clusters, with communication rounds r

Ensure: Global clustering predictions.

for each client $c \in \{1, ..., C\}$ **do**
    Consensus feature initialized by Eq.(1).
end for
for $r = 1$ to R **do**
    for $c = 1$ to C **do**
        Optimize contrastive loss via Eq.(11)
    end for
    Aggregate client models via Eq.(8) and Eq.(9).
    Distribute global model to each client.
end for
Calculate predictions via Eq.(12) and Eq.(13).

Objective Function

The respective total losses for full-views clients f , partial-views clients p and single-view clients s :

$$\mathcal{L} = \begin{cases} \mathcal{L}_f = \mathcal{L}_R^f + \alpha(\mathcal{L}_C^l + \mathcal{L}_C^f) + (1-\alpha)\mathcal{L}_M^f, \\ \mathcal{L}_p = \mathcal{L}_R^p + \alpha\mathcal{L}_C^p + (1-\alpha)\mathcal{L}_M^p, \\ \mathcal{L}_s = \mathcal{L}_R^s + \alpha\mathcal{L}_C^s + (1-\alpha)\mathcal{L}_M^s. \end{cases} \tag{11}$$

Where α denotes a hyperparameter that controls the balance between different objective components. In the optimization of EFDMVC, $\mathcal{L}_R$ serves as the foundation for learning distinct feature representations within individual clients. Concurrently, the cross-view consistency loss $\mathcal{L}_C$ and the model shift loss and $\mathcal{L}_M$ are employed to discover common semantics across views operate synergistically to capture shared semantic patterns across multi-view data.

Finally, the server performs K-means on all common semantics $\mathbf{H}$ to obtain the global centroids $\mathbf{U}$:

$$\min_{\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_K} \sum_{i=1}^{N} \sum_{j=1}^{K} \|\mathbf{H}_i - \mathbf{U}_j\|^2. \tag{12}$$

Therefore, the final prediction result for sample i is:

$$y_i = \arg\min_j \|\mathbf{H}_i - \mathbf{U}_j\|^2. \tag{13}$$

Expriment

Experimental Settings

Scenario Configurations: We implement a privacy-preserving FL framework with the following characteristics: (1) Multi-view data are preprocessed and randomly partitioned into training subsets following *IID* or *Non-IID* distributions; (2) Each client $\mathcal{C}_k$ receives uncertaintyally assigned views $|\mathcal{V}_k| \in [1, V]$ with corresponding data slices; (3) Local models are trained using client-specific multi-view data with encrypted parameter extraction; (4) The central server performs secure aggregation of uploaded parameters via cryptographic protocols; (5) Updated global models are redistributed for subsequent training rounds. The entire process enforces strict *data localization* and *client isolation* principles.

Datasets: We conducted experiments on six datasets: MNIST, HW, Multi-Fashion, BBC (Greene and Cunningham 2006), UCI and Caltech-5V. In order to better simulate real-world heterogeneous scenarios, the data partitioning is done using Dirichlet distribution (Hsu, Qi, and Brown 2019). We configured four heterogeneous environments; Dirichlet (1), Dirichlet (10), Dirichlet (100), and *IID*. Here, smaller values in Dirichlet $(\cdot)$ indicate greater heterogeneity for evaluating the fitness of the EFDMVC and the comparison methods.

Comparison Methods: We select the following five state-of-the-art methods for comparison: MFLVC (Xu et al. 2022), MAGA (Bian et al. 2024), SEM (Xu et al. 2023), HFMVC (Jiang et al. 2024), and FMCSC (Chen et al. 2024). It is worth noting that among the above methods, only HFMVC and FSCSC are applied in a federated environment, whereas the other methods are centralized methods. In order to ensure maximum fairness of comparison, we have made simple modifications to the alignment to support distributed environments. Given the scarcity of FedMVC research, these modifications were unavoidable.

Result Comprison

As systematically summarized in Table 1 , we evaluate EFDMVC against state-of-the-art methods across diverse uncertainty FedMVC scenarios. The comparative analysis across four heterogeneous scenarios reveals four key findings:

(1) EFDMVC demonstrates remarkable clustering accuracy across all experimental configurations, particularly outperforming baseline methods that exhibit catastrophic performance degradation in such uncertainty scenarios. Notably on the Fashion dataset, EFDMVC achieves peak performance margins of 86.89% (ACC), 78.75% (NMI), and 74.95% (ARI) , surpassing existing methods by significant margins.

(2) Compared to traditional centralized algorithms, EFDMVC achieves significant advantages. Furthermore, relative to decentralized algorithms like FMCSC and HFMVC, EFDMVC demonstrates superior performance through : adaptive parameter updating mechanisms, advanced multi-view pattern extraction, and effective resolution of cross-source heterogeneity challenges. Particularly in *IID* scenarios, the clustering performance of HFMVC and FMCSC nearly collapses (with ACC on HW being respectively only 25.60% and 35.70%, while EFDMVC achieves 60.70%), demonstrating the practical feasibility of only EFDMVC.

(3) EFDMVC exhibits superior robustness. For instance, when the degree of heterogeneity ranges from Dirichlet (1.0) to *IID*, EFDMVC's ACC on MNIST ranges from 50.67% to 61.94%, while on UCI, it ranges from 52.05% to 55.45%. In contrast, other methods exhibit much more significant fluctuations. Moreover, as the heterogeneity level decreases, the overall performance of EFDMVC tends to improve. In distributed scenarios, data is divided into more client nodes, resulting in fewer data points for each autoencoder to reconstruct. Conversely, as the heterogeneity level increases,

Data	Heterogeneity	Dirichlet (1.0)			Dirichlet (10)			Dirichlet (100)			*IID*, Dirichlet (∞)		
	Metrics	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
Fashion	MAGA	15.15	0.00	0.00	13.44	0.00	0.00	11.22	0.00	0.00	10.76	0.00	0.00
	MFLVC	45.22	54.53	37.08	69.28	71.45	58.22	58.96	67.74	48.33	62.70	68.60	50.23
	SEM	26.88	25.40	10.04	26.83	24.45	8.82	24.47	24.29	8.51	24.9	25.90	10.67
	HFMVC	28.01	21.96	10.49	24.52	20.89	9.10	21.81	14.07	6.99	23.11	19.29	7.24
	FMCSC	51.93	51.03	36.28	56.82	54.99	41.24	58.21	60.71	46.46	60.06	63.96	50.57
	EFDMVC	**56.60**	**61.87**	**46.00**	**78.03**	**78.01**	**69.50**	**85.16**	**77.82**	**72.88**	**86.89**	**78.75**	**74.95**
MNIST	MAGA	15.50	0.00	0.00	12.45	0.00	0.00	11.76	0.00	0.00	10.89	0.00	0.00
	MFLVC	15.86	8.18	1.34	53.70	44.95	28.38	39.62	31.67	8.10	60.02	47.77	32.65
	SEM	23.19	15.66	7.11	20.01	17.14	6.34	20.86	15.57	6.16	18.94	14.81	4.82
	HFMVC	26.69	20.59	10.21	20.98	10.31	4.03	19.34	9.50	3.40	21.57	17.33	7.71
	FMCSC	41.44	29.95	20.06	58.19	39.81	31.59	55.11	42.15	32.58	53.65	39.54	31.07
	EFDMVC	**50.67**	**36.58**	**27.13**	**61.40**	**45.93**	**38.81**	**61.94**	**48.90**	**42.44**	**75.50**	**57.80**	**54.5**
UCI	MAGA	11.94	0.00	0.00	11.39	0.00	0.00	11.11	0.00	0.00	10.83	0.00	0.00
	MFLVC	28.10	33.91	13.73	30.95	37.60	20.06	34.25	38.38	23.56	36.90	51.96	33.80
	SEM	30.10	25.94	8.58	30.45	27.49	11.55	29.55	30.42	13.11	35.05	31.67	14.35
	HFMVC	29.70	28.56	12.56	23.55	21.17	7.69	23.05	25.02	10.96	27.30	28.36	12.12
	FMCSC	48.90	39.95	27.40	36.15	26.13	14.54	31.50	21.39	12.40	38.40	31.62	18.73
	EFDMVC	**52.05**	**47.28**	**33.13**	**63.60**	**55.11**	**44.08**	**55.95**	**52.19**	**39.08**	**55.45**	**51.02**	**37.09**
HW	MAGA	11.94	0.00	0.00	11.39	0.00	0.00	11.11	0.00	0.00	10.83	0.00	0.00
	MFLVC	34.60	37.18	18.81	37.20	47.46	26.87	34.45	47.33	28.07	36.40	47.63	30.08
	SEM	29.05	32.92	13.02	38.30	38.90	23.40	41.95	43.90	26.81	32.25	36.12	17.18
	HFMVC	31.30	32.83	17.45	30.30	38.64	19.99	30.25	37.12	18.03	26.60	32.93	13.31
	FMCSC	34.60	34.40	18.84	38.20	34.39	19.73	37.05	32.03	19.05	35.70	31.13	16.47
	EFDMVC	**46.80**	**47.38**	**30.18**	**61.95**	**54.00**	**40.26**	**58.55**	**54.23**	**41.02**	**60.70**	**56.92**	**44.47**
Cal-5V	MAGA	20.09	0.00	0.00	17.56	0.00	0.00	19.17	0.00	0.00	20.00	0.00	0.00
	MFLVC	31.64	25.25	14.65	22.07	10.39	2.40	38.36	31.04	**19.19**	34.07	22.57	11.62
	SEM	**39.07**	**27.28**	**16.46**	44.64	34.10	22.59	36.00	**31.31**	18.31	33.29	27.07	14.57
	HFMVC	33.21	22.09	9.98	24.36	8.41	2.82	23.00	9.68	3.19	22.36	8.19	2.71
	FMCSC	34.36	19.92	13.28	30.00	14.58	9.06	31.93	17.79	12.05	44.79	25.73	21.31
	EFDMVC	35.29	18.51	12.94	**52.07**	**39.98**	**30.93**	**42.64**	29.13	18.88	**58.29**	**42.45**	**32.26**

Table 1: Clustering performance. The mean values (%) of 5 runs are reported. The best and the second best values are highlighted in **bold** and underline.

the data reconstructed by individual autoencoders becomes more homogeneous, leading to better results.

Ablation Analysis

We conduct ablation studies on HW and Caltech-5V under both *IID* and *Non-IID* scenarios to evaluate the impact of $\mathcal{L}_C$ (defined in Eq.(4) , Eq.(5), Eq.(6) and Eq.(7)), $\mathcal{L}_M$ (defined in Eq.(8)), and the aggregation strategy (defined in Eq.(9) and Eq.(10)), with results summarized in Table 2 :

(1) Removing $\mathcal{L}_C$ causes the ACC on $\text{HW}_{Dir=1}$ to drop from 49.5% to 27.8% (21.7% decline) and ARI from 32.6% to 9.1%, indicating that the absence of classification loss collapses semantic consistency and fails to capture cross-view clustering structures.

(2) Replacing our aggregation module with FedAvg drastically reduces ACC on $\text{HW}_{Dir=\infty}$ from 69.8% to 25.7%, as FedAvg ignores view quality weights and fails to fuse heterogeneous client knowledge.

(3) Removing $\mathcal{L}_M$ has stronger impacts in *Non-IID* scenarios, e.g., ACC on $\text{Caltech-5V}_{Dir=1}$ decreases by 12.8% (vs. 5.6% in *IID*), due to aggravated local overfitting and failed global knowledge transfer.

(4) The full combination achieves peak performance (e.g., ACC at 69.8% on $\text{HW}_{Dir=\infty}$) through complementary mechanisms: $\mathcal{L}_C$ enforces semantic supervision, $\mathcal{L}_M$ aligns models, and aggregation strategy uncertaintyally fuses heterogeneous knowledge. Overall, the three components act in concert to unleash the model's full potential.

Parameter Analysis

We conducted a detailed analysis of the two hyperparameters through a study on the MNIST dataset, as shown in Fig.3. The study focused on the impact of the loss balancing

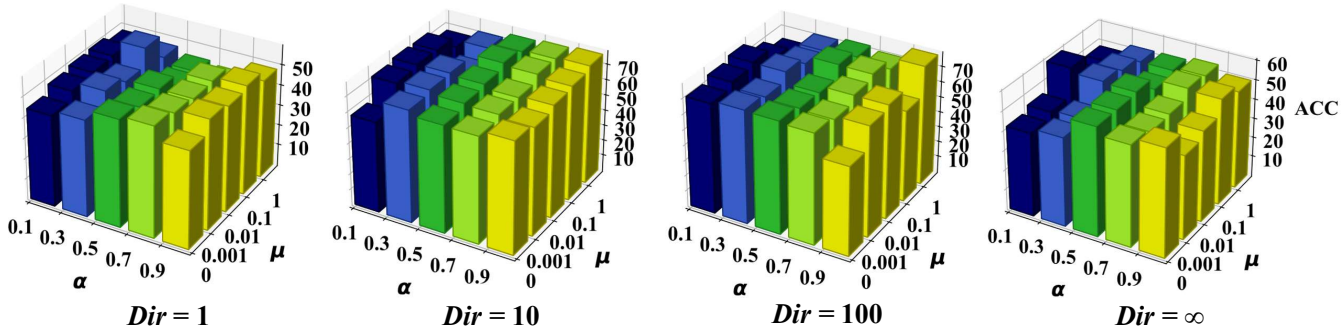

Figure 3: Parameter comprison on different heterogeneous scenarios

Configuration			$\mathbf{HW}_{Dir=1}$			$\mathbf{HW}_{Dir=\infty}$			$\mathbf{Caltech\text{-}5V}_{Dir=1}$			$\mathbf{Caltech\text{-}5V}_{Dir=\infty}$		
$\mathcal{L}_C$	$\mathcal{L}_M$	Agg	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
	✓	✓	27.8	18.8	9.1	28.5	20.9	10.5	29.6	10.2	6.7	34.7	20.1	12.8
✓	✓		27.0	22.8	12.1	25.7	20.6	9.8	23.0	7.3	3.4	34.2	17.0	10.6
✓		✓	46.0	42.5	30.4	64.9	57.0	46.5	34.0	21.4	13.7	54.4	38.3	30.8
✓	✓	✓	**49.5**	**47.0**	**32.6**	**69.8**	**63.0**	**54.5**	**35.8**	**22.4**	**13.7**	**57.0**	**42.3**	**32.0**

Table 2: Ablation studies on both *IID* and *Non-IID* settings

factor α and the regularization coefficient μ. In the experiments, α was varied across $[0.1, 0.3, 0.5, 0.7, 0.9]$, while μ was varied across $[0, 0.001, 0.01, 0.1, 1]$.

The experimental results indicate that when α is set to a small value (e.g., 0.1), the clustering performance of EFDMVC significantly decreases. For instance, under the scenario of α, the ACC metric drops to its lowest value of 44.52%. This may be attributed to the model overly relying on model shift while neglecting the contribution of local models in FL, thereby compromising the ability to extract consistent semantic representations from local multi-view data. In contrast, the regularization coefficient μ exhibits relatively stable influence on the model's performance. Even with varying values of μ, EFDMVC maintains robust performance, indicating that the regularization term plays a crucial role in constraining parameter updates and further emphasizing the importance of the global model in FL frameworks. Considering the experimental results comprehensively, we selected $\alpha = 0.5$ and $\mu = 0.01$ as the default parameters to achieve optimal performance.

Visualization Analysis

We evaluate the clustering accuracy of different models on the Caltech-5V dataset under the *Non-IID* setting, as shown in Fig. 4. Compared with the classical MFLVC, the horizontal FMCSC, and the vertical HFMVC, EFDMVC achieves consistently better performance across all three client models, and its aggregated global model significantly outperforms all baselines. These results demonstrate that the proposed balanced view aggregation module can effectively determine the aggregation direction, whereas existing methods fail to alleviate aggregation uncertainty and often suffer performance degradation after aggregation.Furthermore,

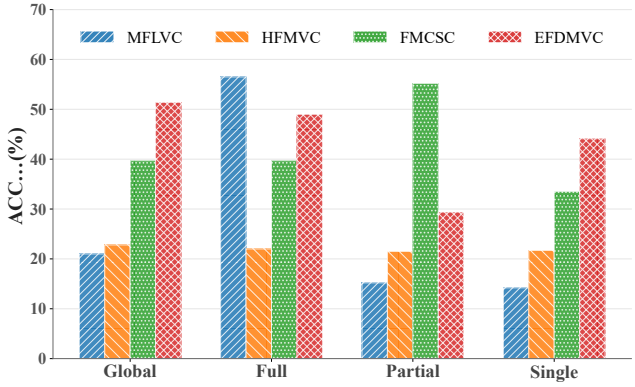

Figure 4: Score comprison on different models

both the full-view and global models achieve better performance, validating our core hypothesis that they should guide the optimization of other models during model drift.

Conclusion

In this paper, we propose EFDMVC, which can handle practical uncertainty scenarios and explore data cluster structures distributed across multiple clients. Extensive experiments demonstrate that EFDMVC outperforms SOTA methods in realistic and broad FedMVC scenarios. Although the current approach still has limitations in explicitly modeling inter-view semantic consistency. Future work will extend this framework to critical applications including uncertainty financial forecasting, while strengthening privacy preservation mechanisms and model lightweight deployment for real-world adaptability.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (No. 62476228), and the Ph.D. Research Project Grant from the Research Institute for Digital Economy and Interdisciplinary Sciences.

References

Bian, J.; Xie, X.; Lai, J.-H.; and Nie, F. 2024. Multi-view contrastive clustering via integrating graph aggregation and confidence enhancement. *Information Fusion*, 108: 102393.

Cao, T.-D.; Truong-Huu, T.; Tran, H.; and Tran, K. 2022. A federated deep learning framework for privacy preservation and communication efficiency. *Journal of Systems Architecture*, 124: 102413.

Che, S.; Kong, Z.; Peng, H.; Sun, L.; Leow, A.; Chen, Y.; and He, L. 2022. Federated Multi-view Learning for Private Medical Data Integration and Analysis. *ACM Transactions on Intelligent Systems and Technology*, 13(4): 1–23.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A Simple Framework for Contrastive Learning of Visual Representations. *arXiv preprint arXiv:2002.05709*.

Chen, X.; Ren, Y.; Xu, J.; Lin, F.; Pu, X.; and Yang, Y. 2024. Bridging gaps: Federated multi-view clustering in heterogeneous hybrid views. *Advances in Neural Information Processing Systems*, 37: 37020–37049.

Chen, X.; Xu, J.; Ren, Y.; Pu, X.; Zhu, C.; Zhu, X.; Hao, Z.; and He, L. 2023. Federated Deep Multi-View Clustering with Global Self-Supervision. In *Proceedings of the 31st ACM International Conference on Multimedia*, 3498–3506.

Fan, T.; Gu, H.; Cao, X.; Chan, C. S.; Chen, Q.; Chen, Y.; Feng, Y.; et al. 2025. Ten Challenging Problems in Federated Foundation Models. *IEEE Transactions on Knowledge and Data Engineering*, 37(7): 4314–4337.

Fang, U.; Li, M.; Li, J.; Gao, L.; Jia, T.; and Zhang, Y. 2023. A Comprehensive Survey on Multi-View Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(12): 12350–12368.

Feng, W.; Liu, D.; Wang, Q.; Liang, W.; and Yan, Z. 2025. Scalable Federated One-Step Multi-View Clustering with Tensorized Regularization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 16586–16594.

Greene, D.; and Cunningham, P. 2006. Practical solutions to the problem of diagonal dominance in kernel document clustering. In *Proceedings of the 23rd international conference on Machine learning*, 377–384.

Grill, J.-B.; Strub, F.; Altché, F.; Tallec, C.; et al. 2020. Bootstrap your own latent: A new approach to self-supervised Learning. arXiv:2006.07733.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum Contrast for Unsupervised Visual Representation Learning. arXiv:1911.05722.

Hinton, G. E.; Krizhevsky, A.; and Wang, S. D. 2011. Transforming auto-encoders. In *International conference on artificial neural networks*, 44–51.

Hsu, T.-M. H.; Qi, H.; and Brown, M. 2019. Measuring the effects of non-identical data distribution for federated visual classification. *arXiv preprint arXiv:1909.06335*.

Huang, S.; Shi, W.; Xu, Z.; Tsang, I. W.; and Lv, J. 2022. Efficient federated multi-view learning. *Pattern Recognition*, 131: 108817.

Huang, W.; Ye, M.; Shi, Z.; Li, H.; and Du, B. 2023. Rethinking federated learning with domain shift: A prototype view. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16312–16322. IEEE.

Jiang, X.; Ma, Z.; Fu, Y.; Liao, Y.; and Zhou, P. 2024. Heterogeneity-Aware Federated Deep Multi-View Clustering towards Diverse Feature Representations. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 9184–9193.

Li, Q.; Diao, Y.; Chen, Q.; and He, B. 2022. Federated learning on non-iid data silos: An experimental study. In *2022 IEEE 38th international conference on data engineering*, 965–978. IEEE.

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*.

Liang, H.; Wei, B.; Cao, X.; Liu, J.; Ouyang, X.; Qiu, J.; Wang, H.; Yang, X.; and Li, T. 2026. Continual deep multi-view clustering via contrastive knowledge replay. *Pattern Recognition*, 172: 112548.

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.

Wu, C.; Wang, H.; Zhang, X.; Fang, Z.; and Bu, J. 2024. Spatio-temporal Heterogeneous Federated Learning for Time Series Classification with Multi-view Orthogonal Training. In *Proceedings of the 32nd ACM International Conference on Multimedia*, 2613–2622.

Xu, J.; Chen, S.; Ren, Y.; Shi, X.; Shen, H.; Niu, G.; and Zhu, X. 2023. Self-Weighted Contrastive Learning among Multiple Views for Mitigating Representation Degeneration. In *Advances in Neural Information Processing Systems*, volume 36, 1119–1131.

Xu, J.; Tang, H.; Ren, Y.; Peng, L.; Zhu, X.; and He, L. 2022. Multi-Level Feature Learning for Contrastive Multi-View Clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16051–16060.

Yang, M.-S.; and Sinaga, K. P. 2025. Federated Multi-View K-Means Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4): 2446–2459.

Yang, X.; Yu, H.; Gao, X.; Wang, H.; Zhang, J.; and Li, T. 2024. Federated Continual Learning via Knowledge Fusion: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, 36(8): 3832–3850.