Graph Consistency and Diversity Measurement for Federated Multi-View Clustering

Bohang Sun[1], Yongjian Deng[1], Yuena Lin[1,2], Qiuru Hai[1], Zhen Yang[1], Gengyu Lyu[1]*

[1]College of Computer Science, Beijing University of Technology
[2]Idealism Beijing Technology Co., Ltd.
sunbohang@emails.bjut.edu.cn, yjdeng@bjut.edu.cn, yuenalin@126.com, haiqiuru@emails.bjut.edu.cn, yangzhen@bjut.edu.cn, lyugengyu@gmail.com

Abstract

Federated Multi-View Clustering (FMVC) aims to learn a global clustering model from heterogeneous data distributed across different devices, where each device only stores one view of all clustering samples. The key to deal with such problem lies in how to effectively fuse these heterogeneous samples while strictly preserve the data privacy across multiple devices. In this paper, we propose a novel structural graph learning framework named MGCD, which leverages both consistency and diversity of multi-view graph structure across global view-fusion server and local view-specific clients to achieve desired clustering while better preserves data privacy. Specifically, in each local client, we design a dual autoencoder to extract the latent consensuses and specificities of each view, where self-representation construction is introduced to generate the corresponding view-specific diversity graph. In the global server, the consistency implied in uploaded diversity graphs are further distilled and then incorporated into the consistency graph for subsequent cross-view contrastive fusion. During the training process, the server generates a global consistency graph and distributes it to each client for assisting in diversity graph construction, while the clients extract view-specific information and upload it to the server for more reliable consistency graph generation. The "server-client" interaction is conducted in an iterative manner, where the consistency implied in each local client is gradually aggregated into the global consistency graph, and the final clustering results are obtained by spectral clustering on the desired global consistency graph. Extensive experiments on various datasets have demonstrated the effectiveness of our proposed method on clustering federated multi-view data.

Introduction

Multi-view data is usually collected from multiple sources, which is represented by several heterogeneous features. For instance, in personal cyber behavioral analysis, diversified individual information are collected from different network platforms and retained in their associated institutions. If we want to give a reliable analysis results, we need to comprehensively consider all personal information held by different institutions. However, in the practical scenarios, these individual sensitive information are prohibited to be exchanged across different institutions since the data privacy is being emphasized. Under such conditions, traditional multi-view learning methods lose their capability to effectively conduct cross-source information fusion and can not comprehensively represent the feature properties of each individual person.

The key to learn from such sensitive multi-view data lies in how to fuse these heterogeneous cross-source features efficiently while preserve the data secrecy of each independent data source. Recently, federated multi-view learning provides an effective solution, which deploys the multi-view clustering method into federated learning framework and fuses these multi-source feature information into a consistency representation for subsequent clustering. For instance, (Huang et al. 2022) proposed a matrix decomposition based method, which orthogonally decomposes the view-specific sample matrix in each client into a basis matrix and a representation matrix and then the representation matrix is uploaded to the sever to mine cross-view consistency for clustering. (Chen et al. 2023) proposed a deep neural network based method, which generates view-specific latent representation by using deep autoencoder in each client and then uploads these representations to the sever to mine cross-view consistency. However, the above federated multi-view clustering methods suffer from some common limitations: (1) All of the above methods leverage cross-view fusion on the latent feature representations uploaded by the clients, which is easily susceptible to model inversion attacks (Sun et al. 2021). (2) These methods only consider cross-view consistencies while the view-specific diversities are regrettably ignored, which naturally results in a suboptimal performance for the final clustering.

To address the above issues, in this paper, we propose a novel structural graph learning framework for multi-view clustering named MGCD, which leverages both consistency and diversity of multi-view graph structure across global view-fusion server and local view-specific clients to achieve desired clustering while better preserves data privacy. Specifically, in each local client, we first design a dual autoencoder to extract the consensus and specific representations of samples respectively, where the two representations can jointly recover the original data to prevent deviation. Afterwards, a unique self-representation reconstruction is introduced to generate diversity graph that represents latent

*Gengyu Lyu is the corresponding author.

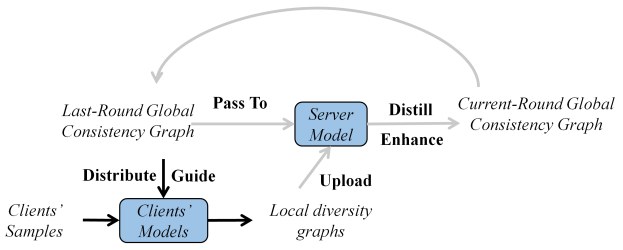


Figure 1: The "server-client" interaction of MGCD.

view-specific diversity relations, where the generation of diversity graph is independent to consensus representation to avoid its negative interference. In the global server, the diversity graphs uploaded from clients are further distilled to refine its implicit consistency information, and these information are incorporated into the last-round global consistency graph for cross-view contrastive fusion to generate the current-round global consistency graph. During the whole training process, different from previous methods that the cross-view consistency extracted by server is only used to guide clients' training, our model pass it to the next-round server's training. As shown in Figure 1, the "server-client" interaction is conducted in an iterative manner like recurrent neural network, where the server model receive last-round global consistency graph and then enhance its consistency by distilling consistency implied in view-specific local diversity graphs, which makes the consistency implied in each local client is gradually aggregated into the global consistency graph and the final clustering results are obtained by spectral clustering based on the final global consistency graph. In summary, the main contributions of our paper lies in the following aspects:

- We propose a novel structured graph learning framework MGCD for federated multi-view clustering, which leverages both consistency and diversity across global view-fusion server and local view-specific clients to achieve desired clustering while preserves data privacy.
- Compared with previous methods, our method is **more secure** and **effective**, which is attributed to the employed safer graph structure to alleviate model inversion attacks and the consistency & diversity measurements to generate more desired global consistency graph.
- Extensive experimental results on various datasets have demonstrated that our proposed model exhibits superior performance against other state-of-the-art algorithms.

Related Work

Multi-view Clustering

Multi-view clustering, unsupervisedly fusing the multi-view data to aid differentiate crucial grouping, is a fundamental task in the fields of data mining (Lyu et al. 2024a,b; Zhong, Lyu, and Yang 2024; Gu et al. 2023; Diallo et al. 2023; Xu et al. 2024; Ma et al. 2024), pattern recognition (Liu et al. 2023; Jiang et al. 2022; Zhang et al. 2023; Liu and Tsang 2017; Hu et al. 2024a; Tao et al. 2022), etc. The key to deal with such problem lies in how to fuse cross-view information and obtain consistent representation for clustering. Current multi-view clustering methods are mainly divided into two categories, i.e., subspace-based methods and graph-based methods. For instance, (Wu, Feng, and Yuan 2024) propose a subspace-based method, which first embeds the multi-view features into a unified kernel tensor and then utilizes the low-rank kernel tensor constraint to capture the consistency information. (Yan et al. 2023) propose a subspace-based method, which uses the autoencoder to learn the latent representation of each view and then introduces constrastive learning to extract cross-view consistent representation. (Wang et al. 2023) propose a graph-based method, which first learns a graph structure of each view by self-representation learning and then generates the consistency graph by fusing these graph structures.

Federated Multi-View Clustering

Federated multi-view clustering aims to cluster multi-view data where the data is distributed among different devices (Huang et al. 2022; Chen and Zhang 2022; Li, Yao, and Liu 2023; Hu et al. 2024b). The key to deal with such problem lies in how to fusion cross-view information under the premise of data privacy. For example, (Ren et al. 2024) propose a self-supervised method, which uses the global consistency prototype from server as self-supervised information to update the latent representation of the sample in each client and then the server combines these representations from clients to update the global consistent prototype. (Chen et al. 2023) propose a deep learning method, which generates view-specific latent representation by deep autoencoder in each client and then the server aligns these representations to mine for consistency representation. Although these methods have made competitive performance in federated multi-view clustering, they still suffer from some drawbacks. (1) Directly sharing the representations of clients are vulnerable to model inversion attacks. (2) These methods only measure cross-view consistency while ignore diversity.

Methodology

In federated multi-view learning, the learning process is generally decomposed into two parts: one global server and multiple local clients. Formally speaking, multi-view data with V views, denoted by $\mathbf{X} = \left\{\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^V\right\}$, are distributed among V different clients. For each client v, its data are represented as $\mathbf{X}^v \in R^{N \times D_v}$, where D_v is the dimensionality of samples in view v and N is the number of samples, v = 1, ...,V. The goal of federated multi-view clustering is to fuse these distributed views in different clients and extract consistency in the global server for subsequent clustering. Notably, during the whole learning process, the data privacy in each client are strictly emphasized and preserved.

Formulation

In this paper, we propose a novel graph-based federated multi-view clustering framework named MGCD, which leverages both consistency and diversity of multi-view graph structure across global server and local clients to achieve

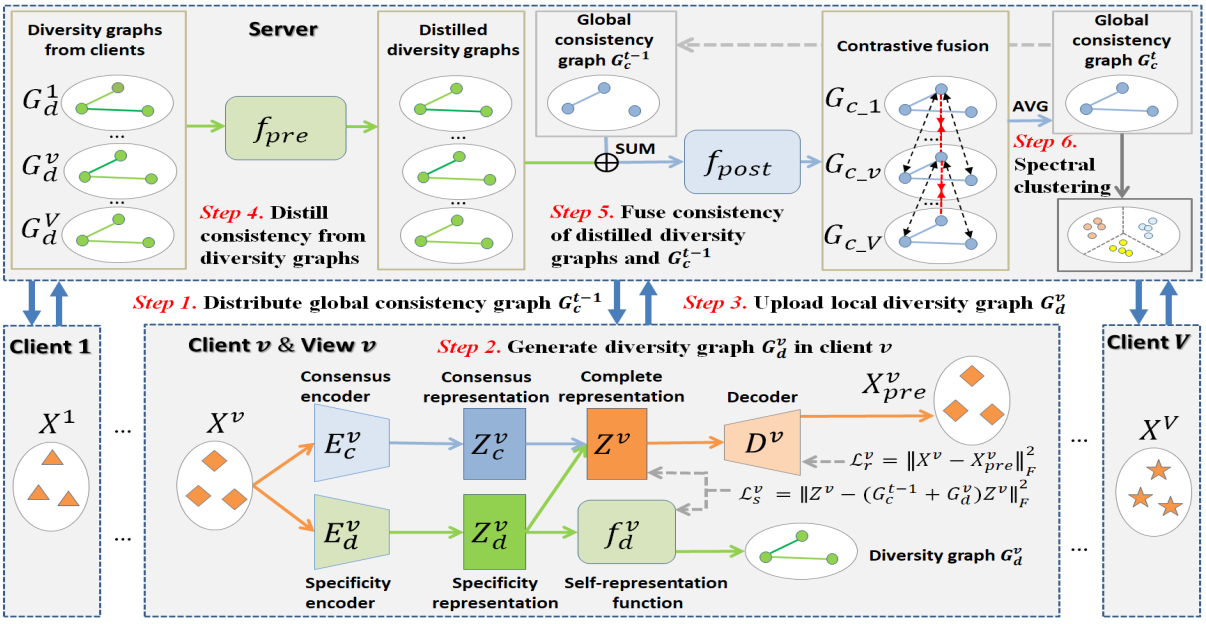


Figure 2: The overview of MGCD, which contains a global server and V local clients. In each global training epoch t, the server firstly distributes global consistency graph G_c^{t-1} to V clients to guide clients' training. Then, each client v generates its corresponding local diversity graph G_d^v in collaboration with G_c^{t-1}. Afterwards, these diversity graphs are sent to the server and further distilled to extract their remaining consistency. Finally, these extracted consistency are incorporated into G_c^{t-1} for contrastive fusion to generate the current global consistency graph G_c^t. After T global training epochs, the desired consistency graph G_c^T is obtained, which is applied to spectral clustering for the final clustering results.

desired clustering while better preserves data privacy. Following previous "server-client" federated architecture, we also decompose our training process into two parts: ***Local Training*** in clients and ***Global Training*** in server. During each training epoch t, the server first receives the last-round global consistency graph $\mathbf{G}_c^{t-1}$ and distributes it to each client. Then, under the guidance of $\mathbf{G}_c^{t-1}$, each client conducts their own *local training* and generates the corresponding local diversity graph $\mathbf{G}_d^v$ to upload to the server. Afterwards, the server conducts its *global training* and fuse cross-view consistency to $\mathbf{G}_c^{t-1}$ to generate the global consistency graph $\mathbf{G}_c^t$. The above operation are conducted in an iterative manner, where the consistent information implied in each local client is gradually aggregated into the global consistency graph, and the final clustering results are obtained by spectral clustering on the final global consistency graph. Figure 2 illustrates the overview of our proposed method.

Local Training In each client v, we separately conduct its own local training to generate the corresponding local diversity graph $\mathbf{G}_d^v$ and upload it to the server for further cross-view fusion. Specifically, we first design a dual autoencoder to extract consensus representation $\mathbf{Z}_c^v \in \mathbb{R}^{N\times d_v}$ and specificity representation $\mathbf{Z}_d^v \in \mathbb{R}^{N\times d_v}$ of $\mathbf{X}^v$ respectively, where the dual autoencoder consists of two dual encoders (consensus encoder $E_c^v(\mathbf{X}^v) : \mathbf{X}^v \mapsto \mathbf{Z}_c^v$ and specificity encoder $E_d^v(\mathbf{X}^v) : \mathbf{X}^v \mapsto \mathbf{Z}_d^v$) and a decoder $D^v(\mathbf{Z}^v) : \mathbf{Z}^v \in \mathbb{R}^{N\times d_v} \mapsto \mathbf{X}_{pre}^v \in \mathbb{R}^{N\times D_v}$. To force the capability of the representations ($\mathbf{Z}_c^v$ and $\mathbf{Z}_d^v$) in recovering complete feature information for each client, a reconstruction loss between the original features $\mathbf{X}^v$ and the corresponding reconstructed features $\mathbf{X}_{pre}^v$ is defined as follows:

$$\begin{aligned}\mathcal{L}_r^v &= \left\|\mathbf{X}^v - D^v\left(\mathbf{Z}^v\right)\right\|_F^2 = \left\|\mathbf{X}^v - D^v\left(\mathbf{Z}_c^v + \mathbf{Z}_d^v\right)\right\|_F^2 \\ &= \left\|\mathbf{X}^v - D^v\left(E_c^v\left(\mathbf{X}^v\right) + E_d^v\left(\mathbf{X}^v\right)\right)\right\|_F^2,\end{aligned} \tag{1}$$

where $\mathbf{Z}^v = \mathbf{Z}_c^v + \mathbf{Z}_d^v \in \mathbb{R}^{N\times d_v}$ indicates the complete representation of $\mathbf{X}^v$, and d_v is the dimensionality of latent representation in v-th client. In traditional federated multi-view learning methods, the above generated representation $\mathbf{Z}^v$ (or $\mathbf{Z}_c^v$ and $\mathbf{Z}_d^v$) is directly uploaded to the server for subsequent cross-view fusion. However, in real-world scenarios, such operation carries **a significant risk of data leakage** since the sever can easily recover the original data by model inversion attack (Sun et al. 2021), especially when the attacker tests out these latent representations come from autoencoder.

To alleviate the model inversion attack and preserve the data security, we intend to bypass traditional feature representation uploading and instead employ the graph structure uploading that only uploads sample relationships to server for subsequent cross-view fusion. Specifically, we define the self-representation term as $\mathbf{Z}^v = \mathbf{G}^v\mathbf{Z}^v$ and utilize a self-representation function $f_d^v(\cdot)$ to generate local diversity graph $\mathbf{G}_d^v$ of each client v with the following loss functions:

$$\begin{aligned}\mathcal{L}_s^v &= \left\|\mathbf{Z}^v - \mathbf{G}^v\mathbf{Z}^v\right\|_F^2 \\ &= \left\|\mathbf{Z}^v - \left(\mathbf{G}_c^{t-1} + \mathbf{G}_d^v\right)\mathbf{Z}^v\right\|_F^2, \\ &s.t.\ \ diag\left(\mathbf{G}^v\right) = 0,\end{aligned} \tag{2}$$

Algorithm 1: The Training Process of MGCD.

Input: Multi-view data $\mathbf{X} = \left\{\mathbf{X}^1, \mathbf{X}^2, \ldots, \mathbf{X}^V\right\}$ distributed in V clients; The number of training epoch T.
Output: Clustering results

Initialize global consistency graph $\mathbf{G}_c^0$ in server;
for epoch $t = 1$ to T
  The clients for v = 1 to V in parallel:
    if $t == 1$ **then**
      Initialize $\{\mathbf{E}_c^v, \mathbf{E}_d^v, \mathbf{D}^v\}$ by minimizing Eq.(1);
    end if
    Receive $\mathbf{G}_c^{t-1}$ from server;
    Local training by Eq.(3) to obtain $\mathbf{G}_d^v$;
    Upload $\mathbf{G}_d^v$ to server;
  The server:
    Receive $\left\{\mathbf{G}_d^1, \mathbf{G}_d^2, ..., \mathbf{G}_d^V\right\}$ from clients;
    Global training by Eq.(6) to obtain $\mathbf{G}_c^t$;
    Distribute $\mathbf{G}_c^t$ to clients;
end for
Perform spectral clustering on $\mathbf{G}_c^T$.

where $\mathbf{G}^v = \mathbf{G}_c^{t-1} + \mathbf{G}_d^v \in \mathbb{R}^{N\times N}$ is the complete graph in v-th client, $\mathbf{G}_c^{t-1} \in \mathbb{R}^{N\times N}$ is the consistency graph distributed from the server, and $f_d^v(\cdot)$ is a fully connected layer. In our method, we employ the global consistency graph $\mathbf{G}_c^{t-1}$ to guide the generation of local diversity graph $\mathbf{G}_d^v$, where $\mathbf{G}_c^{t-1}$ represents the consistency that the server has extracted. Therefore, under the guidance of $\mathbf{G}_c^{t-1}$, the unextracted consistency in $\mathbf{X}^v$ will be first transfered to $\mathbf{G}_d^v$ in client v and then further distilled to generate more compact global consistency graph in the server. The whole loss function in each client is represented as:

$$\mathcal{L}_{client}^v = \mathcal{L}_r^v + \gamma\mathcal{L}_s^v, \tag{3}$$

where γ is a trade-off coefficient between reconstruction loss and self-representation loss. After local training, the generated $\mathbf{G}_d^v$ is uploaded to the server for further global training.

Global Training In the server, we receive the diversity graphs from clients and conduct a unified global training to generate the global consistency graph $\mathbf{G}_c^t$, which is utilized for the final clustering. During the global training process, both the consistency and diversity of multi-view graph structure are measured simultaneously to make $\mathbf{G}_c^t$ more compact so as to improve the final clustering performance. Specifically, considering that the diversity graph $\mathbf{G}_d^v$ uploaded from clients still contains some consistent information, we further distill these diversity graphs and integrate the distilled consistent information into $\mathbf{G}_c^{t-1}$ to generate the refined consistency graph $\mathbf{G}_{c\_v} = f_{post}\left(f_{pre}\left(\mathbf{G}_d^v\right) + \mathbf{G}_c^{t-1}\right) \in R^{N\times N}$, where $f_{pre}(\cdot)$ and $f_{post}(\cdot)$ are two fully connect layers that distill the consistent information from $\mathbf{G}_d^v$ and fuse the distilled consistent information into $\mathbf{G}_c^{t-1}$, respectively.

(1) Graph Consistency Measurement. The above distillation operation generates V refined consistency graphs $\{\mathbf{G}_{c\_v}\}_{v=1}^V$, whose consistency are further measured to realize cross-view fusion. Specifically, we design a graph contrastive fusion strategy, which applies contrastive learning into the refined consistency graphs $\{\mathbf{G}_{c\_v}\}_{v=1}^V$. Such strategy encourages the similarity relationship $\{\mathbf{g}_i^v\}_{v=1}^V$ of each instance across different views to be consistent while requires the similarity relationship of diverse instance to be different, where $\mathbf{g}_i^v$ are the i-th row of the graph $\mathbf{G}_{c\_v}$ and indicates the similarity between the i-th instance and other instances. When formulating the graph contrastive fusion loss, we select $\{(\mathbf{g}_i^v, \mathbf{g}_i^q), v \neq q\}$ to serve as positive pairs and the other relationship pairs to be negative pairs. Accordingly, our designed graph contrastive fusion loss is formulated as:

$$\mathcal{L}_{cl} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{1<v<q<V}\log\frac{e^{s\left(\mathbf{g}_i^v, \mathbf{g}_i^q\right)/\tau_g}}{\sum_{i,j=1\,\frac{i}{\neq}\,j}^{N}\sum_{p=1}^{V}e^{s\left(\mathbf{g}_i^v, \mathbf{g}_j^p\right)/\tau_g}}, \tag{4}$$

where $s(\mathbf{g}_i^v, \mathbf{g}_i^q)$ is similarity between $\mathbf{g}_i^v$ and $\mathbf{g}_i^q$ measured by consine distance, τ_g is a tunable hyper-parameter for the softmax temperature.

(2) Graph Diversity Measurement. While measuring the graph consistency of the refined $\{\mathbf{G}_{c\_v}\}_{v=1}^V$, the specific information in $\mathbf{G}_d^v$ is remained as a refined diversity graph $\mathbf{G}_{c\_d} \in \mathbb{R}^{N\times N}$, where $\mathbf{G}_c^{t-1} + \mathbf{G}_d^v = \mathbf{G}_{c\_v} + \mathbf{G}_{d\_v}$. In real-world scenarios, the $\mathbf{G}_{d\_v}$ tends to contain both view-specific information and noises/outliers in each individual view, which is generally sparse across different views. Thus, we measure the graph diversity by minimizing the sum of the products among $\{\mathbf{G}_{d\_v}\}_{v=1}^V$:

$$\mathcal{L}_{dd} = \sum_{v,q=1\,\frac{v}{\neq}\,q}^{V} Tr\left(\left(\mathbf{G}_{d\_v}\right)\left(\mathbf{G}_{d\_q}\right)^T\right), \tag{5}$$

where $Tr(\cdot)$ represents the trace of a matrix.

By measuring multi-view graph consistency and diversity simultaneously in *Global Training*, we can enforce all consistent information from different clients into the consistency graph $\{\mathbf{G}_{c\_v}\}_{v=1}^V$ and store extra view-specific information or noises/outliers in the diversity graph $\{\mathbf{G}_{d\_v}\}_{v=1}^V$, which can jointly contribute to generate a more compact global consistency graph $\mathbf{G}_c^t = \frac{1}{V}\sum_{v=1}^V \mathbf{G}_{c\_v}$. The whole loss function in the global server is represented as:

$$\mathcal{L}_{server} = \mathcal{L}_{cl} + \lambda\mathcal{L}_{dd}. \tag{6}$$

Optimization

Algorithm 1 summarizes the training process of MGCD, which consists of two main parts: the clients and the server, where the clients perform local training in parallel and the server conducts global training for multi-view fusion. At each training epoch t, each client first receives the global consistency graph $\mathbf{G}_c^{t-1}$ and conducts local training to generate its local diversity graph $\mathbf{G}_d^v$. Then, these local diversity graphs are uploaded to the server for global training, which generates the global consistency graph $\mathbf{G}_c^t$ and distribute it to clients for the next round training. After T round iterations, we obtain the desired global consistency graph $\mathbf{G}_c^T$, which can perform spectral clustering on it for clustering results.

The Superiorities of MGCD

1) Our proposed MGCD is more secure. In local client, we upload the graph structure to the server instead of the auto-encoder feature representations, which can effectively alleviate the model inversion attack and reduce the risk of data leakage. Especially, our graph structure only contains instance similarity relationship while not contains any latent feature representations, which can prevent the attacker from recovering the original data easily even if he has tested out the data generation strategy in clients.

2) Our proposed MGCD is more effective. In global server, we leverages both consistency and diversity of multi-view graph structure to generate more compact global consistency graph for clustering. Compared with previous methods that only measure consistency, such operation can gradually aggregate the consistency information from different clients into the global consistency graph while store the remaining view-specific information and noises/outliers in the corresponding diversity graphs.

The Further Explanation of MGCD

1) What is the diversity in MGCD. In our paper, the diversity represents a much broader concept than view-specific feature attributes. It could be caused by not only view-specific attributes, but also noise and outliers, which is not conducive to clustering. Additionally, diversity measurement emphasizes the cross-view mutual exclusions in diversity graphs, formulated by Eq. (5), which will lead the learned consistency in $\{\mathbf{G}_{c\_v}\}_{v=1}^{V}$ to be more compact.

2) Why distill consistency from diversity graphs. At training epoch t, the global consistency graph $\mathbf{G}_c^{t-1}$ represents the consistency that the server has extracted. However, at the beginning of model training, there must be some unextracted consistency that implied in each client. Besides, in local training, the diversity graphs are generated under the guidance of $\mathbf{G}_c^{t-1}$, which results in the unextracted consistency is transferred to diversity graphs from clients' samples. Therefore, it is necessary to distill consistency from diversity graphs. Finally, as the server continues to distill, the diversity graphs will eventually contain only diversity.

Experiments

Experimental Settings

Datasets We employ six widely-used multi-view datasets for comparative studies, including *Mfeat* (Wang, Yang, and Liu 2019), *Scene* (Fei-Fei and Perona 2005), *Aloi* (Li et al. 2023), *Animal* (Li et al. 2016), *Cifar10* (Zhang et al. 2018) and *NoisyMNIST* (Peng et al. 2019). The specific characteristics of these datasets are recorded in Table 1.

Data	Samples	Clusters	View dimensions
Mfeat	2000	10	216/76/64/6/240/47
Scene	4485	15	20/59/40
Aloi	10800	100	77/13/64/125
Animal	11673	20	2689/2000/2001/2000
Cifar10	50000	10	512/2048/1024
NoisyMNIST	50000	10	784/784

Table 1: Statistical characteristics of the six datasets.

The Compared Methods In order to verify the effectiveness of our proposed MGCD, we employ eight state-of-the-art multi-view clustering methods for comparative experiments, which includes five centralized methods of **LMVSC** (Kang et al. 2020), **SiMVC** (Trosten et al. 2021), **CoMVC** (Trosten et al. 2021), **MFLVC** (Xu et al. 2022), **GCFAgg** (Yan et al. 2023) and three federated methods of **FedMVL** (Huang et al. 2022), **FedDMVC** (Chen et al. 2023), **FCUIF** (Ren et al. 2024), where all compared methods are implemented according to the source codes released by the authors, and the optimal parameters are set according to the suggestion in the corresponding literature.

Metrics There are four widely-used metrics applied to quantitatively evaluate the performance of multi-view clustering methods, including Accuracy (ACC), Normalized Mutual Information (NMI), Purity(Pur) and Adjusted Rand Index (ARI), whose detailed definitions are illustrated in (Liang et al. 2022). For each of the above metric, the higher value indicates the better performance.

Implementation Details. The diversity encoder E_d^v, consistency encoder E_c^v and decoder D^v are formulated by four fully-connected layers and the dimensions are set to $\{D_v, 500, 500, 2000, 512\}$, $\{D_v, 500, 500, 2000, 512\}$ and $\{512, 2000, 500, 500, D_v\}$ respectively, where the activation function is RELU. The $f_d^v(\cdot)$ in client is composed of a fully-connected layer with dimendisons $\{512, N\}$. To reduce the size of server model, the dimensions of $f_{pre}(\cdot)$ and $f_{post}(\cdot)$ are respectively set as $\{N, 500, N\}$ and $\{N, 500, N\}$, where $N >> 500$. At the first training epoch, we pre-train dual autoencoder 20 epochs in each client. Then, at the following training epoch, the clients and server iteratively train on mini-batches of size 256 by using Adam optimizer(Kingma and Ba 2014) with learning rate of 0.000001 in PyTorch(Paszke et al. 2019) framework. The hyperparameters γ and λ are set to 100 and 0.001 respectively. All experiments are conducted on the same machine with the Intel(R) Xeon(R) Gold 6148 2.40GHz CPU, GeForce RTX 3090 GPUs, and 512GB RAM.

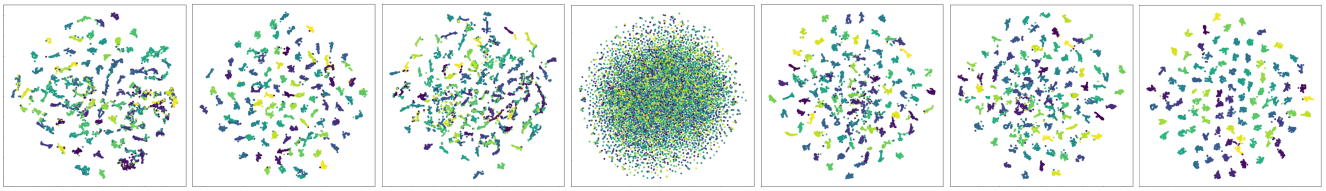
Experimental Results

Comparisons with other methods Table 2 records the experimental comparisons between our proposed MGCD and the other 8 comparing methods, where the best and the sub-optimal performance are highlighted in bold and underlined, respectively. In addition, Figure 3 illustrates the visualization of clustering results of each method on the *Aloi* dataset. According to Table 2 and Figure 3, we can observe that:

(1) Among the employed all datasets, our MGCD is superior to all comparing methods on all evaluation metrics, even has a significant leading gap compared with sub-optimal methods. Especially on the *Animal* dataset, the improvements over the sub-optimal method are 32.91%, 48.31%, 32.91%, and 36.85% on ACC, NMI, ARI and PUR, respectively. These experimental results demonstrate the effectiveness of our proposed method and we attribute such success

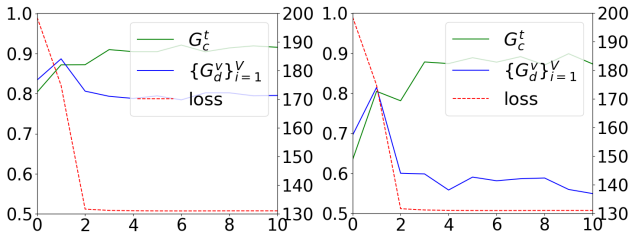
Data set	Metric	Centralized					Federated			
		LMVSC	SiMVC	CoMVC	MFLVC	GCFAgg	FedMVL	FedDMVC	FCUIF	Ours
Mfeat	ACC	0.6550	0.8001	0.7750	0.8665	0.5945	0.1300	0.9365	0.9341	**0.9390**
	NMI	0.6386	0.8407	0.8243	0.8736	0.7401	0.0085	0.9063	0.8951	**0.9173**
	ARI	0.5283	0.7563	0.7207	0.8166	0.5043	0.1790	0.9002	0.8854	**0.9005**
	PUR	0.7461	0.8413	0.8147	0.8665	0.6565	0.1320	0.9503	0.9491	**0.9671**
Scene	ACC	0.3222	0.4383	0.4347	0.3173	0.2022	0.0945	0.4360	0.4252	**0.5285**
	NMI	0.3396	0.4657	0.4627	0.3392	0.1842	0.0100	0.4184	0.3880	**0.5625**
	ARI	0.1714	0.2787	0.2710	0.1784	0.0746	0.0643	0.2697	0.2474	**0.3665**
	PUR	0.3922	0.5084	0.5001	0.3456	0.2486	0.1064	0.4237	0.4020	**0.5845**
Aloi	ACC	0.6390	0.6730	0.7010	0.7490	0.5745	0.0349	0.8566	0.7460	**0.9160**
	NMI	0.7700	0.8530	0.8940	0.8570	0.8268	0.0731	0.9210	0.8426	**0.9668**
	ARI	0.5030	0.5550	0.6530	0.6680	0.5184	0.0374	0.8050	0.6285	**0.8992**
	PUR	0.6900	0.8150	0.7940	0.7810	0.5968	0.0361	0.8941	0.7926	**0.9302**
Animal	ACC	0.1310	0.1600	0.1560	0.1910	0.1528	0.0791	0.1829	0.1793	**0.5201**
	NMI	0.0290	0.1360	0.1350	0.1660	0.1480	0.0147	0.1641	0.1590	**0.6491**
	ARI	0.0290	0.0530	0.0500	0.0750	0.0639	0.0125	0.0694	0.0665	**0.4010**
	PUR	0.1390	0.1720	0.1640	0.2030	0.1929	0.1010	0.1720	0.1673	**0.5715**
Cifar10	ACC	0.8753	0.8359	0.9275	0.9925	0.9902	0.1646	0.9917	0.9888	**0.9948**
	NMI	0.7798	0.7324	0.8925	0.9795	0.9744	0.0533	0.9781	0.9719	**0.9851**
	ARI	0.3274	0.8057	0.9836	0.9836	0.9787	0.1673	0.9818	0.9756	**0.9885**
	PUR	0.8753	0.8359	0.9275	0.9925	0.9902	0.1691	0.9917	0.9888	**0.9948**
NoisyMNIST	ACC	0.3274	0.3831	0.4141	0.2497	0.6465	0.1246	0.4564	0.4263	**0.7791**
	NMI	0.3027	0.3266	0.4047	0.2054	0.6469	0.0130	0.4137	0.3893	**0.7261**
	ARI	0.1603	0.2988	0.3616	0.0778	0.4522	0.1221	0.2861	0.2758	**0.6351**
	PUR	0.5196	0.4109	0.4667	0.1905	0.6497	0.1467	0.4564	0.4263	**0.7791**

Table 2: Comparative results between our proposed MGCD and 8 state-of-the-art methods on six datasets, where the best results are presented in bold and the second-best are in underline.



(a) GCFAGG (b) MFLVC (c) LMVSC (d) FedMVL (e) FedDMVC (f) FCUIF (g) Ours

Figure 3: The visualizations of the clustering results of different methods on *Aloi* dataset.



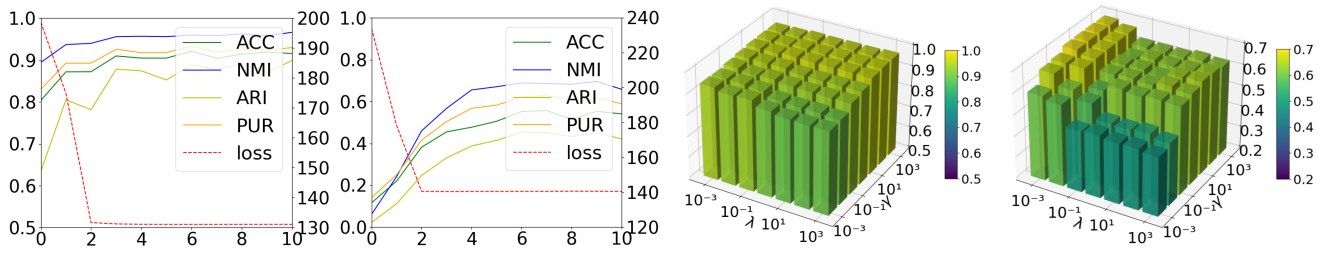
(a) ACC (b) ARI

Figure 4: The clustering performance comparisons between the consistency graph $\mathbf{G}_c^t$ and the diversity graph $\{\mathbf{G}_d^v\}_{v=1}^V$ on *Aloi* datasets as the training epoch increases.

to the measurement of consistency and diversity in server, which makes our model generate a more compact consistency graph with a clear cluster structure.

(2) As shown in Figure 3, we select three centralized multi-view consistency clustering methods (GCFAGG, MFLVC, LMVSC) and three federated multi-view consistency clustering methods (FedMVL, FedDMVC, FCUIF) to conduct the visualization comparisons of clustering results with our proposed MGCD. We can observe that our MGCD exhibits a more clear cluster structure than all other methods, which demonstrates the superiority of MGCD in exploring consistency information across different views.

Comparisons between global consistency graph and local diversity graphs of our model Figure 4 illustrates the clustering performance comparisons between the consistency graph $\mathbf{G}_c^t$ and the diversity graphs $\{\mathbf{G}_d^v\}_{i=1}^V$ on *Aloi* dataset, where the diversity graph is represented by the averaging of local diversity graphs from V clients. According to Figure 4, we can find that, as the training epoch increases, the clustering performance of consistency graph gradually increases while that of diversity graph gradually decreases, which demonstrates that our proposed MGCD gradually transfer the consistent information from local diversity graphs to the global consistency graph.



(a) Aloi (b) Animal (c) Aloi (d) Animal

Figure 5: The convergence analysis and parameter analysis on *Aloi*, *Animal* datasets respectively.

	$\mathcal{L}_s$	$\mathcal{L}_r$	$\mathcal{L}_{cl}$	$\mathcal{L}_{dd}$	ACC	NMI	ARI	PUR
(A)	✓		✓	✓	0.853	0.924	0.774	0.868
(B)	✓	✓		✓	0.903	0.954	0.867	0.919
(C)	✓	✓	✓		0.807	0.899	0.599	0.838
(D)	✓	✓			0.795	0.883	0.598	0.816
(E)	✓	✓	✓	✓	**0.916**	**0.966**	**0.899**	**0.930**

Table 3: Ablation studies of loss components on *Aloi* dataset.

	Dual AE	Distillation	ACC	NMI	ARI	PUR
(a)		✓	0.796	0.898	0.638	0.822
(b)	✓		0.859	0.927	0.777	0.874
(c)	✓	✓	**0.916**	**0.966**	**0.899**	**0.930**

Table 4: Ablation studies on some components of MGCD.

Model Analysis

Ablation Study. We conduct two series of ablation studies from the perspective of loss functions and model components. Table 3 records the loss ablation studies on *Aloi* dataset, where $\mathcal{L}_s$ is the loss to generate diversity graph in client, $\mathcal{L}_r$ is the loss to obtain latent representation of samples in client, $\mathcal{L}_{cl}$ is the loss to measure consistency in server and $\mathcal{L}_{dd}$ is the loss to measure diversity in server. Table 4 records the model ablation studies on *Aloi* dataset, where *Dual AE* represents dual autoencoder in client and *Distillation* represents the consistency distillation operation $f_{pre}(\cdot)$ in server. According to Table 3-4, we can find that:

(1) According to Table 3, (E) is superior to (A), which indicates that the reliable representation of samples is helpful to generate desired graphs and can further contribute to improve the clustering performance. Meanwhile, (E) also shows better clustering performance than (B), (C) and (D), which demonstrates the effectiveness of our employed graph consistency and diversity measurement. In addition, according to the comparison between (B) and (C), we can find that both graph consistency measurement and graph diversity measurement are helpful to the generation of global consistency graph, while the later has greater contribution.

(2) In Table 4, (a) replaces the designed dual autoencoder with traditional autoencoder in each client and (b) removes the consistency distillation operation $f_{pre}(\cdot)$ from global training in server. According to Table 4, (c) shows better performance than (a), which indicates that the specificity representation from our designed dual autoencoder is more suitable for the generation of diversity graph in each client and our designed dual autoencoder performs significant superiorities against traditional autoencoder. Meanwhile, (c) outperforms (b), which demonstrates the fact that the diversity graphs from clients still contain consistent information. Our designed distillation operation can gradually transfer the consistent information implied in diversity graph to the global consistency graph, which avoids the loss of consistent information from local clients and enhances the robustness of the global consistency graph.

Convergence analysis. Figure 5 shows the convergence curves of MGCD on *Aloi*, *Animal* datasets, where the values of loss and evaluation metrics are illustrated in each subfigure. According to Figure 5, we can observe that the value of loss drops significantly at the beginning of the iteration process and gradually reaches stability as the number of iterations increases. And the values of evaluation metrics gradually increase and fluctuate in a narrow range. These results verified the convergence of our proposed MGCD.

Parameter sensitivity analysis. We experimentally evaluate the effect of hyperparameters on the clustering performance of MGCD, which includes γ in clients and λ in server. Figure 5 shows the NMI metric value of MGCD on *Aloi*, *Animal* datasets, where γ is varied from 10^{-3} to 10^{3} and λ from 10^{-3} to 10^{3}. According to Figure 5, the clustering results of MGCD are insensitive to both γ and λ ranging from 10 to 1000, and 0.001 to 0.01, respectively. In our experiments, we set γ to 100 and λ to 0.001.

Conclusion

In this paper, we proposed a new graph-based federated multi-view clustering method, which leverages both consistency and diversity of multi-view graph structure to achieve desired clustering while preserves data privacy. Compared with previous methods, our proposed method conducts multi-view fusion according to the graph structure rather than feature representation, which can better mitigate model inversion attacks and preserve the client privacy. Meanwhile, the measurement of both consistency and diversity can further assist in generating more compact consistency graph, which naturally improves the final clustering performance. Extensive experimental results on various datasets have verified the effectiveness of our proposed method.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (No. 2023YFB3107100), the National Natural Science Foundation of China (No. 62306020, 62203024, 62173286), the Young Elite Scientist Sponsorship Program by BAST (No. BYESS2024199), the R&D Program of Beijing Municipal Education Commission (No. KM202310005027), the Major Research Plan of National Natural Science Foundation of China (No. 92167102), and the Beijing Natural Science Foundation (No. L244009).

References

Chen, J.; and Zhang, A. 2022. Fedmsplit: Correlation-adaptive federated multi-task learning across multimodal split networks. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 87–96.

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 *ACM International Conference on Multimedia*, 3498–3506.

Diallo, B.; Hu, J.; Li, T.; Khan, G. A.; Liang, X.; and Wang, H. 2023. Auto-attention mechanism for multi-view deep embedding clustering. *Pattern Recognition*, 143: 109764.

Fei-Fei, L.; and Perona, P. 2005. A bayesian hierarchical model for learning natural scene categories. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 524–531.

Gu, Z.; Feng, S.; Hu, R.; and Lyu, G. 2023. ONION: Joint Unsupervised Feature Selection and Robust Subspace Extraction for Graph-based Multi-View Clustering. *ACM Transactions on Knowledge Discovery from Data*, 17(5): 1–23.

Hu, D.; Liu, S.; Wang, J.; Zhang, J.; Wang, S.; Hu, X.; Zhu, X.; Tang, C.; and Liu, X. 2024a. Reliable Attribute-missing Multi-view Clustering with Instance-level and feature-level Cooperative Imputation. In *ACM International Conference on Multimedia*, 1456–1466.

Hu, X.; Qin, J.; Shen, Y.; Pedrycz, W.; Liu, X.; and Liu, J. 2024b. An Efficient Federated Multiview Fuzzy C-Means Clustering Method. *IEEE Transactions on Fuzzy Systems*, 32(4): 1886–1899.

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

Jiang, B.; Xiang, J.; Wu, X.; Wang, Y.; Chen, H.; Cao, W.; and Sheng, W. 2022. Robust Multi-View Learning via Adaptive Regression. *Information Sciences*, 610: 916–937.

Kang, Z.; Zhou, W.; Zhao, Z.; Shao, J.; Han, M.; and Xu, Z. 2020. Large-scale multi-view subspace clustering in linear time. In *AAAI Conference on Artificial Intelligence*, 4412–4419.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 1–15.

Li, S.; Yao, D.; and Liu, J. 2023. Fedvs: Straggler-resilient and privacy-preserving vertical federated learning for split models. In *International Conference on Machine Learning*, 20296–20311.

Li, X.; Ren, Z.; Sun, Q.; and Xu, Z. 2023. Auto-weighted tensor schatten p-norm for robust multi-view graph clustering. *Pattern Recognition*, 134: 109083.

Li, Y.; Shi, X.; Du, C.; Liu, Y.; and Wen, Y. 2016. Manifold regularized multi-view feature selection for social image annotation. *Neurocomputing*, 204(5): 135–141.

Liang, Y.; Huang, D.; Wang, C. D.; and Philip, S. Y. 2022. Multi-view graph learning by joint modeling of consistency and inconsistency. *IEEE Transactions on Neural Networks and Learning Systems*, 35(2): 2848–2862.

Liu, W.; and Tsang, I. W. 2017. Making Decision Trees Feasible in Ultrahigh Feature and Label Dimensions. *Journal of Machine Learning Research*, 18(81): 1–36.

Liu, W.; Yuan, J.; Lyu, G.; and Feng, S. 2023. Label driven latent subspace learning for multi-view multi-label classification. *Applied Intelligence*, 53(4): 3850–3863.

Lyu, G.; Kang, W.; Wang, H.; Li, Z.; Yang, Z.; and Feng, S. 2024a. Common-Individual Semantic Fusion for Multi-View Multi-Label Learning. In *International Joint Conference on Artificial Intelligence*, 4715–4723.

Lyu, G.; Yang, Z.; Deng, X.; and Feng, S. 2024b. L-VSM: Label-Driven View-Specific Fusion for Multiview Multilabel Classification. *IEEE Transactions on Neural Networks and Learning Systems*, 1–15.

Ma, H.; Wang, S.; Yu, S.; Liu, S.; Huang, J.; Wu, H.; Liu, X.; and Zhu, E. 2024. Automatic and Aligned Anchor Learning Strategy for Multi-View Clustering. In *ACM International Conference on Multimedia*, 5045–5054.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; and Chintala, S. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, 1–12.

Peng, X.; Huang, Z.; Lv, J.; Zhu, H.; and Zhou, J. T. 2019. COMIC: Multi-view clustering without parameter selection. In *International Conference on Machine Learning*, 5092–5101.

Ren, Y.; Chen, X.; Xu, J.; Pu, J.; Huang, Y.; Pu, X.; Zhu, C.; Zhu, X.; Hao, Z.; and He, L. 2024. A Novel Federated Multi-View Clustering Method For Unaligned And Incomplete Data Fusion. *Information Fusion*, 108: 102357.

Sun, J.; Li, A.; Wang, B.; Yang, H.; Li, H.; and Chen, Y. 2021. Soteria: Provable defense against privacy leakage in federated learning from representation perspective. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9311–9319.

Tao, L.; Feng, L.; Wei, H.; Yi, J.; Huang, S. J.; and Chen, S. 2022. Can Adversarial Training Be Manipulated By Non-Robust Features? In *Advances in Neural Information Processing Systems*, 26504–26518.

Trosten, D. J.; Lokse, S.; Jenssen, R.; and Kampffmeyer, M. 2021. Reconsidering representation alignment for multi-view clustering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1255–1265.

Wang, H.; Yang, Y.; and Liu, B. 2019. GMC: Graph-based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 32(6): 1116–1129.

Wang, J.; Feng, S.; Lyu, G.; and Gu, Z. 2023. Triple-Granularity Contrastive Learning for Deep Multi-View Subspace Clustering. In *ACM International Conference on Multimedia*, 2994–3002.

Wu, T.; Feng, S.; and Yuan, J. 2024. Low-Rank Kernel Tensor Learning for Incomplete Multi-View Clustering. In *AAAI Conference on Artificial Intelligence*, 15952–15960.

Xu, J.; Ren, Y.; Wang, X.; Feng, L.; Zhang, Z.; Niu, G.; and Zhu, X. 2024. Investigating and Mitigating the Side Effects of Noisy Views for Self-Supervised Clustering Algorithms in Practical Multi-View Scenarios. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 22957–22966.

Xu, J.; Tang, H.; Ren, Y.; Peng, L.; Zhu, X.; and He, L. 2022. Multi-level feature learning for contrastive multi-view clustering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16051–16060.

Yan, W.; Zhang, Y.; Lv, C.; Tang, C.; Yue, G.; Liao, L.; and Lin, W. 2023. GCFAgg: Global and Cross-view Feature Aggregation for Multi-view Clustering. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 19863–19872.

Zhang, C.; Jiang, B.; Wang, Z.; Yang, J.; Lu, Y.; Wu, X.; and Sheng, W. 2023. Efficient Multi-View Semi-Supervised Feature Selection. *Information Sciences*, 649: 119675.

Zhang, Z.; Liu, L.; Shen, F.; Shen, H. T.; and Shao, L. 2018. Binary multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(7): 1774–1782.

Zhong, Q.; Lyu, G.; and Yang, Z. 2024. Align While Fusion: A Generalized Nonaligned Multiview Multilabel Classification Method. *IEEE Transactions on Neural Networks and Learning Systems*, 1–10.