



ARTICLE

Multi-View Deep Fuzzy Clustering for Data Representation Learning

Jianing Zhang¹, Zhikui Chen^{1,*}, Jing Gao¹ and Peng Li²

¹School of Software, Dalian University of Technology, Dalian, China

²School of Computer Science and Technology, Dalian University of Technology, Dalian, China

*Corresponding Author: Zhikui Chen. Email: zkchen@dlut.edu.cn

Received: 25 November 2025; Accepted: 03 March 2026; Published: 08 May 2026

ABSTRACT: With the increasing development of ocean information technology, the multi-view fuzzy clustering is attracting increasing attention in pattern mining for massive multi-view ocean data of heterogeneous distributions, owing to its superior performance. However, the previous multi-view fuzzy clustering methods cannot fully consider informative topologies hidden in data distributions, which are crucial to recognize partitions of data. Moreover, they fail to capture invariant structures of multi-view ocean data in learning clustering-specific fusion representation. In addition, they do not take into consideration consistencies contained in the manifolds of data generation in mining soft patterns. To address those challenges, the deep multi-view generative fuzzy contrastive clustering (DMGFCC) is proposed within a Siamese architecture, which captures soft patterns of data via clustering-specific fusion representations of invariant structures in informative topologies. To be specific, a multi-view Siamese generative adversarial architecture is designed to capture the joint distribution of data as well as invariant structures, which is composed of the view-specific generator network providing pairwise implicit constraints, the view-specific discriminator network distilling knowledge of real data, and the view-specific cluster network capturing fuzzy patterns of fusion information. Furthermore, a generative adversarial dual contrastive clustering loss is devised, which consists of a generative adversarial loss fitting data distributions and a dual contrastive clustering loss learning soft patterns with consistencies of data manifolds. Finally, extensive experiments are conducted on four benchmark datasets, and the results demonstrate the competitive performance compared with the 11 representative methods.

KEYWORDS: Neural networks; fuzzy systems; pattern mining; multi-view feature fusion

1 Introduction

With the increasing development of ocean information technology, the information from the ocean has demonstrated significant potential across a wide range of fields, such as ocean exploration and ocean current prediction [1–4]. Large amounts of data are inevitably produced and collected from ocean sensors, which are often classified as multi-view data. For example, underwater equipment may collect text, image, sound, and video information to derive meaningful insights into marine environments. These multi-view data describe a richer picture of the real world by consistent and complementary knowledge in heterogeneous views. Thus, they pose great challenges to the effective mining of hidden patterns, which becomes a crucial task in the future ocean computational intelligence.

Multi-view fuzzy clustering, as a fundamental technique of unsupervised learning, captures robust patterns of multi-view data from the consistent and complementary knowledge of heterogeneous views [5,6]. It extracts fuzzy knowledge from heterogeneous data by softening crisp partition boundaries to possibilistic memberships that measure similarities of data. Early multi-view fuzzy clustering algorithms were based

on fuzzy c -means clustering and possibilistic c -means, which can be approximately split into collaborative multi-view fuzzy clustering (Co-MvFCM) and multi-kernel multi-view fuzzy clustering (Mk-MvFCM). Co-MvFCM mines consensus patterns via mutual links of intra-view soft partitions rooted in fuzzy prototype clustering schemes. For instance, Jiang et al. [7] utilized soft partitions to capture view-specific data structures and mine view-common fuzzy patterns via inter-view mutual links with entropy maximizing. Mk-MvFCM utilizes the collaborative learning of intra-view local structures of the kernel space to explore common patterns of views. For example, Zeng et al. [8] integrated the view-specific local partition with the view-common global partition to mine consensus fuzzy patterns on the basis of a common latent space of multiple kernels. Those early multi-view fuzzy clustering methods usually perform in the shallow feature spaces, which cannot well capture intrinsic patterns hidden in inter-view nonlinear correlations and intra-view deep semantics.

Recently, deep multi-view clustering methods have attracted much attention, which leverage hierarchical nonlinear transformations to merge consistent and complementary knowledge of inter-view correlations and intra-view semantics for pattern mining of multi-view data. For example, deep canonical correlation analysis utilizes the maximization of correlation between views to mine a view-common feature subspace with complementary information for clustering pattern recognition [9]. Deep multi-view subspace clustering learns a data-driven self-expression coefficient matrix to integrate complementary information between views, which uses the linear dependence constraint to uncover the subspace of each class [10]. Deep multi-view matrix factorization discovers a set of subspace bases with the non-negative constraint to capture a consensus feature space, and then leverages k -means to learn clustering patterns [11]. Deep multi-view spectral clustering learns a common eigenvector matrix of views by reducing the discrepancy between the common eigenvector matrix and private eigenvector matrices [12].

Although those edge-cutting deep multi-view clustering methods have achieved promising performance, most of them mine data patterns on the basis of local structure information rooted in point-to-point mappings of data reconstructions, instead of informative topologies hidden in data distributions that are crucial to recognize partitions of data. Furthermore, the deep multi-view clustering methods usually utilize the single-flow computing architecture which cannot ensure invariant structures of data in clustering-specific fusion representation learning of consistent and complementary knowledge in views. In addition, they do not take into consideration consistencies contained in the manifolds of data generation in an unsupervised manner. Thus, the deep multi-view clustering methods cannot well fit the joint distribution in data of complex distributions.

To address those challenges of pattern mining in ocean multi-view data with complex heterogeneous distribution, the deep multi-view generative fuzzy contrastive clustering (DMGFCC) method is proposed, which captures fuzzy patterns from complementary knowledge hidden in multi-view data as well as clustering-specific fusion representations in an end-to-end manner. Specifically, a multi-view Siamese generative adversarial architecture is designed with two symmetrical sister networks, which can explicitly capture data-invariant structures of multi-view data in clustering-specific fusion representation learning.

In the Siamese architecture, each sister network, which learns distributions of data generation and data partition in an end-to-end paradigm, is composed of the view-specific generator network, the view-specific discriminator network, and the view-specific cluster network. The view-specific generator network fits a probabilistic generation function between noisy inputs of category-static information and observable samples, which produces multi-view data with pair-wise constraints. The view-specific discriminator network models a conditional discriminant function to provide supervision information for joint distribution fitting of multi-view data via distinguishing real data from generated data. The view-specific cluster network utilizes pair-wise indicator features to measure the implicit constraints of multi-view data. Furthermore, a generative

adversarial dual contrastive clustering loss composed of a generative adversarial loss and a dual contrastive clustering loss is introduced. The generative adversarial loss assists the view-specific generator network and the view-specific discriminator network to produce data with category knowledge that are subject to the real distribution. The dual contrastive clustering loss helps the view-specific generator network and the view-specific cluster network to capture fuzzy patterns of multi-view data, which can encourage consistencies of data manifolds in the generative clustering with implicit pair-wise knowledge.

The main contributions of this paper are listed as follows:

- The deep multi-view generative fuzzy contrastive clustering (DMGFCC) is designed within a Siamese architecture composed of the view-specific generator network, the view-specific discriminator network, and the view-specific cluster network, which can capture fuzzy informative patterns as well as clustering-specific fusion representations.
- The generative adversarial dual contrastive clustering loss is devised to train model parameters, which consists of a generative adversarial loss fitting the joint distributions between data and categories and a dual contrastive clustering loss capturing fuzzy patterns of multi-view data by fusion representations.
- Extensive experiments are conducted on four benchmark datasets to validate model performance. The results illustrate that DMGFCC achieves competitive performance compared with compared representative methods, especially on complex large datasets.

The rest of this paper is organized as follows: [Section 2](#) gives a detailed review of related work. [Section 3](#) describes the proposed multi-view Siamese generative adversarial architecture. [Section 4](#) introduces the generative adversarial dual contrastive clustering loss, and [Section 5](#) presents the extensive results. Finally, [Section 6](#) concludes this work.

2 Related Work

2.1 Shallow Multi-View Fuzzy Clustering

Existing shallow multi-view fuzzy clustering methods can be roughly classified into two categories, i.e., collaborative multi-view fuzzy clustering (Co-MvFCM) and multi-kernel multi-view fuzzy clustering (Mk-MvFCM).

Co-MvFCM mines fuzzy patterns of multi-view data via knowledge transfer of view-specific complementary information. For example, Cleuziou et al. [5] conduct fuzzy *c*-means clustering on views and reduce the inter-view disagreements between view-specific membership degrees to infer the stable consensus partition of data, i.e., the geometric mean of view-specific membership degrees. Jiang et al. [7] design the view weights measuring the importance of views in the convergence process of fuzzy clustering to focus on the contribution of key views in generating consensus data partition. Wang and Chen [13] leverage min-max mechanism with variable weights of views indicating the current view of the highest cost to perform fuzzy clustering on each view in turn to construct consensus patterns with low costs. Yang and Sinaga [14] construct the view-level and feature-level weight scheme via discovering the core feature in key view in fuzzy clustering, to mine consensus fuzzy patterns in a focused manner. Zhang et al. [15] design cross-view anchor graph indicating the similarity relationship of multi-view data for latent information learning to directly obtain the clustering result in one-step clustering. Yin et al. [16] utilize local structure preserving mechanism to balance the global and local information during discriminative clustering, where the intra-cluster compactness and inter-cluster separability are considered simultaneously.

Mk-MvFCM captures soft patterns of multi-view data via utilizing the nonlinear kernel to compress mismatches between data manifolds and distance metrics in similarity measurements. For instance, Tzortzis and Likas [17] design the linear combination of multiple kernels to extract representations of multi-view data

on which the fuzzy clustering patterns are mined. Guo et al. [18] propose the cooperation fuzzy clustering scheme which alternatively performs the multi-kernel fusion and pattern recognition to model consensus clustering partition. Ye et al. [19] utilize a co-regularized kernel fuzzy clustering algorithm to mine the consensus partition of multi-view data via maximizing adaptive similarities between view-common and view-specific clustering indicators. Zeng et al. [8] leverage the multikernel framework to fit the mapping from data space to a common kernel space via endowing each view with multiple kernels, where fuzzy clustering of views is conducted to produce the global partition.

2.2 Deep Multi-View Fuzzy Clustering

Deep multi-view fuzzy clustering introduces various deep neural networks into the pattern mining of fuzzy clustering, which enables the fitting of latent complex distributions. Trosten et al. [20] deploy deep networks with clustering and contrastive heads to extract the view-weighted deep representations of multi-view data for predicting fuzzy clustering partitions, where contrastive learning is utilized to improve the separation between representations. Gao et al. [21] leverage the DCCA (deep canonical correlation analysis) autoencoder architecture to enhance the view-common self-expression matrix via the canonical correlation maximization between deep representations of views, where the self-expression matrix is utilized in spectral clustering to produce fuzzy clustering partitions. Mao et al. [22] design mutual information-based fuzzy clustering networks via implementing the inter-view mutual information maximization and intra-view mutual information minimization to capture view-common class information and reduce view-specific details, respectively, which can devise an unadulterated consensus partition of multi-view data. Cheng et al. [23] construct the multi-view graph convolution networks to induce consensus clustering assignment via graph relationships into embedding representations for fuzzy clustering. Yin et al. [24] propose a variational autoencoder combined with a Gaussian mixture prior distribution via modeling the generation process of multi-view data to predict the conditional posterior distributions of clusters. Shi et al. [25] utilize an entropy regularized self-weighted autoencoder with consensus membership to tune the membership uniformity and reduce cluster assignment discrepancy among views of data, for consistent fuzzy clustering partitions.

3 Deep Multi-View Generative Fuzzy Contrastive Clustering

The deep multi-view generative fuzzy contrastive clustering (DMGFCC) is devised within a Siamese architecture that utilizes invariant structures and informative topologies of data distributions to mine soft patterns. Furthermore, it learns clustering-specific fusion representations of multi-view data via contrastive learning of consistent and complementary knowledge, which can guarantee the intra-cluster compactness and the inter-cluster separation of data manifolds in pattern mining. As shown in Fig. 1, DMGFCC is constructed as dual computing flows of symmetrical sister networks that are composed of the view-specific generator network, the view-specific discriminator network, and the view-specific cluster network. The view-specific generator network fits the joint distribution between category variables and sample variables to capture informative topologies for data generation with implicit constraints. The view-specific discriminator network assists the view-specific generator network to fit real distributions of data via the adversarial game theory between real data and fake data. The view-specific cluster network utilizes the inter-view and intra-view contrastive learning of category-invariant structures to explore fuzzy patterns hidden in data manifolds as well as clustering-specific representations.

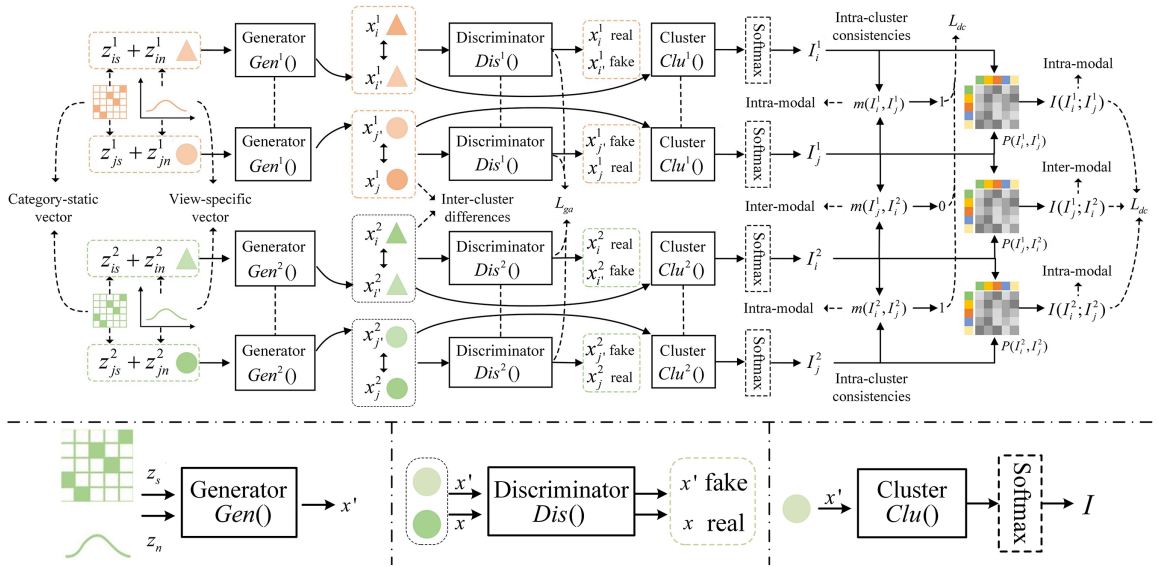


Figure 1: The architecture of DMGFCC. Top: Overall architecture. Bottom-Left: The view-specific generator network. Bottom-Center: The view-specific discriminator network. Bottom-Right: The view-specific cluster network.

3.1 The View-Specific Generator Network

The view-specific generator network learns a generative mapping function of data joint distributions, which is responsible for information transfer between the class space and the data space in each view. It captures knowledge of informative topologies via synthesizing pair-wise multi-view data with the help of intra-cluster invariant structures and view-specific perturbations. To be specific, the view-specific generator network of the v -th view $Gen^v()$ is modeled as the following form:

$$x_{gi}^v = Gen^v(z_{gi}^v; \theta_g^v), z_{gi}^v = z_{in}^v + z_{is}^v \quad (1)$$

where x_{gi}^v is the i -th synthetic sample in the v -th view, and $Gen^v()$ denotes the generator of the deep neural network with the parameters θ_g^v in the v -th view. z_{gi}^v represents the noise vector to trigger data generation in which z_{is}^v and z_{in}^v are the category-static vector preserving the intra-cluster invariance and the view-specific vector ensuring complementary knowledge of views, respectively.

In the Siamese network, the view-specific generator network uses the generative mapping function of the joint distribution of data generation to generate pairwise multi-view data belonging to one out of the predefined K categories, which can deceive the view-specific discriminator networks. For instance, if fed with two sets of hidden vectors with the same category-static vectors, the view-specific generator network synthesizes data belonging to the same clusters. Otherwise, it produces data of different clusters. Consequently, the generator network can promote the intra-cluster consistencies and inter-cluster inconsistencies of data manifolds in the view-specific cluster networks.

3.2 The View-Specific Discriminator Network

The view-specific discriminator network aims to model a statistical decision function that can accurately distinguish synthetic multi-view data with real multi-view data, providing supervision information for the view-specific generator network in joint distribution fitting. The view-specific discriminator network $Dis^v()$ in the v -th view is computed in the following form:

$$f_i^v = Dis^v(x_i^v; \theta_d^v), x_i^v \sim P(x^v) \text{ or } x_i^v \sim P(x_g^v | z_g^v) \quad (2)$$

where $f_i^v \in [0, 1]$ is the probability to measure the authenticity of data in the v -th view, and $Dis^v()$ denotes the discriminator of the deep neural network with the parameters θ_d^v . $P(x^v)$ and $P(x_g^v | z_g^v)$ represent the distributions of real data and fake data in the v -th view, respectively.

In the Siamese architecture, the view-specific discriminator network utilizes the adversarial game of structure information hidden in synthetic data and real data, assisting the view-specific generator network to produce multi-view data with pairwise constraints in joint distribution fitting. For instance, if the view-specific discriminator network captures divergencies of intrinsic structures between synthetic data and real data, i.e., a low probability output of the view-specific discriminator networks (Eq. (2)) for fake data, the synthetic data cannot follow the intrinsic distribution of real data. In other words, there are significant divergences between synthetic data and real data. Then, the view-specific generator network utilizes that divergency information of data structures to enhance the quality of fake data of predefined categories by optimizing parameters towards deceiving the view-specific discriminator network.

3.3 The View-Specific Cluster Network

The view-specific cluster network learns a semantic mapping from the data space to the pattern space, where samples are grouped into each cluster via soft memberships. It captures fuzzy patterns of complementary information in multi-view data, utilizing indicator features in which each element denotes the membership probability. That is, data are transformed into vectors with the dimension being the number of pre-defined clusters, where each dimension denotes a kind of pattern. To this end, the indicator feature is stacked at the end of the view-specific cluster network, computed via:

$$I_i^v = Clu^v(x_i^v; \theta_c^v) \quad (3)$$

where I_i^v is the indicator feature for the i -th sample in the v -th view. $Clu^v()$ is the v -th cluster network of the deep neural network with parameters θ_c^v . Furthermore, I_i^v follows the constraint:

$$\text{Cont} - I : \|I_i^v\|_1 = 1, I_{ik}^v \geq 0, k = 1, 2, \dots, K \quad (4)$$

in which I_{ik}^v is the k -th element of I_i^v . $I_{ik}^v \in [0, 1]$ denotes the partition probability that the i -th sample belongs to the k -th cluster. $\|I_i^v\|_1$ is the L1-norm of I_i^v , i.e., the sum of all I_{ik}^v .

Afterwards, the view-specific cluster network learns the fusion representations of multi-view data via the average of indicator features in each view, which puts constraints on mining of consensus patterns. The fusion representation is computed via:

$$I_{fi} = \sum_v^V I_i^v / V \quad (5)$$

In the Siamese architecture, the view-specific cluster network utilizes pairwise data with implicit constraints to extract clustering-specific fusion representations of inter-view complementary information as well as fuzzy consensus patterns of intra-cluster invariant structures. For instance, data of the same clusters are input into the view-specific cluster networks; the indicator features activate in the same elements. Otherwise, indicator features activate different elements. By optimizing the implicit constraints, the view-specific cluster network ensures inter-view complementarity and consistency of multi-view data.

4 The Generative Adversarial Dual Contrastive Clustering Loss

In this section, a generative adversarial dual contrastive clustering loss (GADCCL) is designed to supervise the learning of DMGFCC. It utilizes invariant semantics hidden in consistent and complementary information of multi-view data to learn fuzzy patterns, as well as clustering-specific fusion representations. Moreover, GADCCL leverages invariant structures of data endowed by the inter-view and intra-view Siamese architectures to preserve consistencies of manifolds in multi-view data fuzzy clustering. GADCCL is composed of a generative adversarial loss fitting joint distributions between classes and samples in views and a dual contrastive loss mining fuzzy multi-view consensus patterns of views. GADCCL is computed as follows:

$$L_{gadc} = \lambda L_{ga} + \gamma L_{dc} \quad (6)$$

where λ and γ are the trade-off hyper-parameters for the generative adversarial loss L_{ga} and the dual contrastive loss L_{dc} , balancing the influence between joint distribution fitting and fuzzy pattern mining.

4.1 The Generative Adversarial Loss

The generative adversarial loss L_{ga} guides the generator in each view to model the joint distributions of data with the help of the discriminator, leveraging knowledge from the adversarial games between real samples and fake samples. It helps DMGFCC to capture invariant semantics and structures via endowing data with pairwise cluster constraints. L_{ga} is expressed in the following form:

$$L_{ga} = \min \max \sum_{v=1}^V \mathbb{E}_{x^v \sim P(x^v)} (Dis^v(x^v)) + \mathbb{E}_{x^v \sim P(x_g^v | z_g^v)} (1 - Dis^v(x^v)) \quad (7)$$

where \mathbb{E} is the expectation operation. $P(x^v)$ and $P(x_g^v | z_g^v)$ are the distributions of real data and fake data in the v -th view, respectively.

The generative adversarial loss ensures that the deep fuzzy multi-view Siamese network produces inter-view and intra-view pairwise data. That is, it endows data with pair-wise implicit knowledge which facilitates the intra-cluster consistencies in fuzzy pattern mining, as well as for inter-view complementarities in representation learning for the view-specific cluster networks.

4.2 The Dual Contrastive Loss

The dual contrastive loss L_{dc} assists the Siamese cluster networks to mine intrinsic patterns and the clustering-specific fusion representations with the help of the implicit invariant structures endowed by DMGFCC. That is, L_{dc} guides the cluster networks to capture a transformation function between the original data space \mathbb{R}^D and the latent feature space \mathbb{R}^K , such that samples belonging to the same cluster are closer than samples of different clusters in the latent feature space \mathbb{R}^K . L_{dc} is defined as the dual contrastive learning with fusion clustering, which is composed of the instance contrastive learning, the distribution contractive learning, and the fusion clustering. The instance contrastive learning is defined as a max-min loss of sample similarities to capture local invariant structures and semantics. The distribution contractive learning is devised as the mutual information loss of indicator features to distill global structures and semantics. The fusion clustering aggregates complementary information from view-specific clustering representations to further guide learning of the Siamese cluster networks.

$$L_{dc} = L_{cl} + L_{mi} + L_{fc} \quad (8)$$

4.2.1 The Max-Min Contrastive Learning Loss

The max-min contrastive learning loss is designed based on the triplet form of contrastive learning. It is able to excavate structures hidden in the inter-view and intra-view consistent and complementary information in the local instance perspective. To be specific, given the multi-view dataset $X = \{x_i^v\}_{i=1, v=1}^{N, V}$, $N = N_1, \dots, N_K$, the cluster network $Clu^v()$, the max-min contrastive learning loss is computed as follows:

$$L_{cl} = \sum_{v=1}^V \sum_{k=1}^K \sum_{i=1}^{N_k} (\|Clu_{s1}^v(x_{i,k}^v) - Clu_{s2}^v(x_{j,k}^v)\|_2 - \|Clu_{s1}^v(x_{i,k}^v) - Clu_{s2}^v(x_{l,h}^v)\|_2 + \alpha) \quad (9)$$

$$+ \sum_{v=1}^V \sum_{w \neq v}^V \sum_{k=1}^K \sum_{i=1}^{N_k} (\|Clu^v(x_{i,k}^v) - Clu^w(x_{i,k}^w)\|_2 - \|Clu^v(x_{i,k}^v) - Clu^w(x_{l,h}^w)\|_2 + \alpha)$$

where $\|\cdot\|_2$ is the L2-norm to represent the concept distance of samples and α is the marginal constant distance to promote the compactness of clusters. $Clu_{s1}^v()$ and $Clu_{s2}^v()$ are two sister networks in the v -th view. $Clu^v()$ and $Clu^w()$ are one of the two sister view-specific cluster networks in the v -th view and the w -th view, respectively. $x_{i,k}^v$ ($* = i, j$) denotes a sample in the k -th class of the v -th view. $x_{l,h}^v$ represents the l -th sample in the h -th class ($h \neq k$). In the max-min contrastive learning loss, the first term utilizes the contrastive learning of intra-view invariant structures to capture fuzzy patterns by virtue of indicator features with the implicit pairwise constraints. The second term employs contrastive inter-view consistent and complementary semantics to further promote the mining of fuzzy patterns on the basis of clustering-specific representations.

To accurately measure similarities of samples in fuzzy pattern mining of the view-specific cluster networks, the concept distance between indicator features of samples, outputs of the view-specific cluster networks, is computed by the cosine similarity as follows:

$$m(I_i, I_j) = \frac{\langle Clu(x_i), Clu(x_j) \rangle}{\|Clu(x_i)\|_2 \cdot \|Clu(x_j)\|_2} \quad (10)$$

where $\langle \cdot, \cdot \rangle$ is the dot product of vectors.

Afterwards, the contrastive loss is re-computed in the following form:

$$L_{cl} = \sum_{v=1}^V \sum_{k=1}^K \sum_{i=1}^{N_k} (m(I_{i,k}^v, I_{l,h}^v) - m(I_{i,k}^v, I_{j,k}^v) + \alpha) + \sum_{v=1}^V \sum_{w \neq v}^V \sum_{k=1}^K \sum_{i=1}^{N_k} (m(I_{i,k}^v, I_{l,h}^w) - m(I_{i,k}^v, I_{i,k}^w) + \alpha) \quad (11)$$

At the same time, each indicator feature I^v follows the constraint Cont-I (Eq. (4)), representing the soft partition of samples. $m(I_i, I_j) \in [0, 1]$ and each $m(I_i, I_j)$ can be interpreted as the probability that the i -th sample and the j -th sample belong to the same cluster. Thus, L_{cl} can be recast as the max-mix game:

$$L_{cl} = \min \max \sum_{v=1}^V (\mathbb{E}_{I_i, I_j \sim P(c_i)} (1 - m(I_{i,k}^v, I_{j,k}^v)) + \mathbb{E}_{I_i \sim P(c_i), I_j \sim P(c_j)} m(I_{i,k}^v, I_{j,k}^v)) \quad (12)$$

$$+ \min \max \sum_{v=1}^V (\mathbb{E}_{I_i, I_j \sim P(c_i)} (1 - m(I_{i,k}^v, I_{j,k}^w)) + \mathbb{E}_{I_i \sim P(c_i), I_j \sim P(c_j)} m(I_{i,k}^v, I_{j,k}^w))$$

where samples of the same cluster maximize the probability $m(I_i, I_j)$ and samples of different clusters minimize the probability.

4.2.2 The Mutual Information Contrastive Learning Loss

The mutual information contrastive learning loss is derived from the entropy form of contrastive learning. It utilizes the implicit pairwise constraints to maximize the mutual information of intra-view invariant structures and inter-view consistent and complementary information in the global distribution perspective. Specifically, given the multi-view dataset $X = \{x_i^v\}_{i=1, v=1}^{N, V}$, $N = N_1, \dots, N_K$, the cluster network $Clu^v(\cdot)$, the mutual information contrastive learning loss is computed as follows:

$$L_{mi} = - \sum_{v=1}^V \sum_{k=1}^K \sum_{i, j=1}^{N_k} \log \frac{\exp(m(I_{i,k}^v, I_{j,k}^v)/\tau)}{\sum_{j=1}^N \exp(m(I_{i,k}^v, I_{j,k}^v)/\tau)} - \sum_{v=1}^V \sum_{t \neq v}^V \sum_{k=1}^K \sum_{i, j=1}^{N_k} \log \frac{\exp(m(I_{i,k}^v, I_{j,k}^t)/\tau)}{\sum_{j=1}^N \exp(m(I_{i,k}^v, I_{j,k}^t)/\tau)} \quad (13)$$

where τ is a temperature parameter, and $\exp(\cdot)$ denotes the natural exponential function. In the mutual information contrastive learning loss, the first term maximizes dependency of intra-view invariant structures to explore intra-cluster consistencies and inter-cluster differences. The second term utilizes dependency of consistent and complementary information to enhance clustering-specific representation learning in soft partitions.

In the mutual information contrastive learning loss, $m(I_i, I_j)$ denotes the cosine distance between the i -th sample and the j -th sample, measuring similarities of pairwise samples. It can be interpreted as the joint probability of co-occurrence in the same cluster. Thus, the conditional probability of the i -th sample and the j -th sample can be re-defined as:

$$Q(I_{j,l}^v | I_{i,k}^v) = \frac{\exp(m(I_{i,k}^v, I_{j,l}^v)/\tau)}{\sum_{l'=1}^K \sum_{h=1}^{N_{l'}} \exp(m(I_{i,k}^v, I_{h,l'}^v)/\tau)} \quad (14)$$

At the same time, the true condition probability of the i -th sample and the j -th sample, derived from the implicit pairwise constraints, is expressed as follows:

$$P(I_{j,l}^v | I_{i,k}^v) = \begin{cases} 1, & I_{j,l}^v \in C_v(I_{i,k}^v) \\ 0, & I_{j,l}^v \in \bar{C}_v(I_{i,k}^v) \end{cases} \quad (15)$$

where $C_v(I_{i,k}^v)$ is the cluster of $I_{i,k}^v$ with N_k samples, and $\bar{C}_v(I_{i,k}^v)$ is the complementary set of $C_v(I_{i,k}^v)$.

Thus, the v -th intra-view mutual information contrastive loss is defined via:

$$\begin{aligned} L_v^{itra} &= -\mathbb{E}_{P(I_{i,k}^v)} \left(\sum_{j=1}^{N_k} \log \frac{\exp(m(I_{i,k}^v, I_{j,k}^v)/\tau)}{\sum_{l=1}^K \sum_{h=1}^{N_l} \exp(m(I_{i,k}^v, I_{h,l}^v)/\tau)} \right) \\ &= -\mathbb{E}_{P(I_{i,k}^v)} \left(\sum_{j=1}^{N_k} \log Q(I_{j,k}^v | I_{i,k}^v) \right) \\ &= -\mathbb{E}_{P(I_{i,k}^v)} \left(\sum_{j=1}^N P(I_{j,k}^v | I_{i,k}^v) \log Q(I_{j,k}^v | I_{i,k}^v) \right) \\ &\geq -\mathbb{E}_{P(I_{i,k}^v)} \left(\sum_{j=1}^N P(I_{j,k}^v | I_{i,k}^v) \log P(I_{j,k}^v | I_{i,k}^v) \right) \\ &= H(I_{s2}^v) - I(I_{s1}^v; I_{s2}^v) \end{aligned} \quad (16)$$

where $H(\cdot)$ is the entropy function and $I(\cdot; \cdot)$ is the mutual information function. The intra-view mutual information contrastive loss maximizes the dependency of intra-view invariant structures to encourage the intra-cluster compactness and inter-cluster separation.

Similarly, the inter-view mutual information contrastive loss between the ν -th view and the t -th view is computed via:

$$\begin{aligned}
 L_{\nu t}^{iter} &= -\mathbb{E}_{P(I_{i,k}^{\nu})} \left(\sum_{j=1}^{N_k} \log \frac{\exp(m(I_{i,k}^{\nu}, I_{j,k}^t)/\tau)}{\sum_{l=1}^K \sum_{h=1}^{N_l} \exp(m(I_{i,k}^{\nu}, I_{h,l}^t)/\tau)} \right) \\
 &= -\mathbb{E}_{P(I_{i,k}^{\nu})} \left(\sum_{j=1}^{N_k} \log Q(I_{j,k}^t | I_{i,k}^{\nu}) \right) \\
 &= -\mathbb{E}_{P(I_{i,k}^{\nu})} \left(\sum_{j=1}^N P(I_{j,k}^t | I_{i,k}^{\nu}) \log Q(I_{j,k}^t | I_{i,k}^{\nu}) \right) \\
 &\geq -\mathbb{E}_{P(I_{i,k}^{\nu})} \left(\sum_{j=1}^N P(I_{j,k}^t | I_{i,k}^{\nu}) \log P(I_{j,k}^t | I_{i,k}^{\nu}) \right) \\
 &= H(I_{s2}^t) - I(I_{s1}^{\nu}; I_{s2}^t)
 \end{aligned} \tag{17}$$

The inter-view mutual information contrastive loss promotes the fusion of consistent and complementary information in soft pattern mining.

Thus, mutual information contrastive learning loss is summed as follows:

$$L_{mi} = \sum_{\nu} L_{\nu}^{itra} + \sum_{\nu} \sum_{t \neq \nu} L_{\nu t}^{iter} \tag{18}$$

4.2.3 The Fusion Clustering Loss

The fusion clustering loss aligns complementary information from view-specific clustering representations to further guide learning of the Siamese cluster networks. To be specific, given the view-specific fusion indicator feature $I^{\nu} \in \mathbb{R}^K$ with $\nu = 1, 2, \dots, V$, the fusion clustering loss is computed as follows:

$$L_{fc} = \sum_{\nu=1}^V D_{KL}(P || Q^{\nu}) = \sum_{\nu=1}^V \sum_{i=1}^N \sum_{j=1}^K p_{ij} \log \frac{p_{ij}}{q_{ij}^{\nu}} \tag{19}$$

where Q^{ν} is the clustering-inference distributions of the ν -th view, and P is the auxiliary target distribution. Each Q^{ν} is obtained via the Student's t-distribution:

$$q_{ij}^{\nu} = \frac{(1 + \|I_i^{\nu} - \mu_j^{\nu}\|_2^2)^{-1}}{\sum_j (1 + \|I_i^{\nu} - \mu_j^{\nu}\|_2^2)^{-1}} \tag{20}$$

in which μ_j denotes the j -th prototype initialized by k-means. P is defined as the mean of Q^{ν} :

$$p_{ij} = \sum_{\nu} q_{ij}^{\nu} / V \tag{21}$$

where p_{ij} denotes the mean assignment of the i -th sample on the basis of Eq. (20).

The details of DMGFCC are outlined in Algorithm 1.

Algorithm 1: DMGFCC

Input: Input multi-view dataset $X = \{x_i^v\}_{i=1, v=1}^{N, V}$, predefined clusters K , and training epochs n .

Randomly initialize parameters $\theta = \{\theta_g^v, \theta_d^v, \theta_c^v\}_{v=1}^V$.

for $t = 1, \dots, n$ **do**

Sample pairwise input vectors $z_{g_i}^v$ and $z_{g_j}^v$ for each view.

Generate pairwise instances of views:

$$x_{g_i}^v = \text{Gen}^v(z_{g_i}^v; \theta_g^v), x_{g_j}^v = \text{Gen}^v(z_{g_j}^v; \theta_g^v).$$

Distinguish real data and generated data:

$$f_i^v = \text{Dis}^v(x_i^v; \theta_d^v), x_i^v \sim P(x^v) \text{ or } x_i^v \sim P(x_g^v | z_g^v),$$

$$f_j^v = \text{Dis}^v(x_j^v; \theta_d^v), x_j^v \sim P(x^v) \text{ or } x_j^v \sim P(x_g^v | z_g^v).$$

Calculate clustering indicator feature in each view:

$$I_i^v = \text{Clu}^v(x_{g_i}^v; \theta_c^v), I_j^v = \text{Clu}^v(x_{g_j}^v; \theta_c^v).$$

Calculate the fusion representation:

$$I_{fi} = \sum_v^V I_i^v / V, I_{fj} = \sum_v^V I_j^v / V.$$

Obtain the clustering assignment:

$$c_i = \arg \max_{k \leq K} I_{fi, k}, c_j = \arg \max_{k \leq K} I_{fj, k}.$$

Calculate the adversarial clustering loss L .

Update discriminator network parameters θ_d^v by $\nabla_{\theta_d^v} L$.

Update cluster network parameters θ_c^v by $\nabla_{\theta_c^v} L$.

Update generator network parameters θ_g^v by $\nabla_{\theta_g^v} L$.

end for

for $x_i \in X$ **do**

$$I_i^v = \text{Clu}^v(x_i^v; \theta_c^v).$$

$$I_{fi} = \sum_v^V I_i^v / V.$$

$$c_i = \arg \max_{k \leq K} I_{fi, k}.$$

end for

Output: Clustering assignment $\{c_i\}_{i=1}^N$.

5 Experiments

In this section, extensive experiments are conducted on four multi-view benchmark datasets to validate the performance of DMGFCC, compared with 11 representative methods. All the experiments are implemented by Python.

5.1 Evaluation Datasets

Four multi-view benchmark datasets are utilized to assess the performance of all the 11 clustering methods. The statistics are listed in [Table 1](#) with the following descriptions.

- **MNIST-USPS**, a benchmark multi-view image dataset, is composed of 7291 samples where two views of the similar distributions are extracted from MNIST (Modified National Institute of Standards and Technology) and USPS (United States Postal Service), respectively.
- **MNIST-EDGE**, a benchmark multi-view image dataset, consists of 54,000 samples. It is of high complexity in the volume and variety. Each sample utilizes the original digital image and the edge digital image as two views.

- **MNIST-INVERSE** is a two-view image dataset composed of 60,000 samples, where each sample is represented by the original digital image and the digital image with inverse pixels. The dataset is also with a complex distribution.
- **EDGE-INVERSE** is composed of 54,000 samples with two views, in which the edge digital image and the inverse digital image are used as views to represent samples. This dataset is more complex than other three datasets in the distributions.

Table 1: The details of four datasets in the experiments.

Dataset	Numbers	Clusters	Dim-View 1	Dim-View 2
MNIST-USPS	7291	10	784	784
MNIST-EDGE	54,000	10	784	784
MNIST-INVERSE	60,000	10	784	784
EDGE-INVERSE	54,000	10	784	784

5.2 Clustering Metrics

Five clustering metrics are used to fully validate the performance of DMGFCC. For the clustering metrics, the high value indicates the good performance, and the detail definitions are listed as follows:

- **Accuracy (ACC)** is defined as the ratio of the number of samples with correct assignment to the number of samples, by comparing clustering-inference assignment with ground-truth assignment in the following form:

$$ACC = \frac{1}{N} \sum_{i=1}^N \delta(g_i, \text{map}(c_i)) \quad (22)$$

where N is the number of samples, $\delta()$ denotes an indicator function, and $\text{map}()$ is implemented by the Hungarian algorithm. c_i and g_i are the clustering-inference label and the ground-truth label of the i -th sample, respectively.

- **Normalized Mutual Information (NMI)** measures correlations between clustering-inference assignment and ground-truth assignment via the entropy theory, defined as follows:

$$NMI = \frac{2MI(C, G)}{H(C) + H(G)} \quad (23)$$

where C and G denote the clustering-inference assignment and the ground-truth assignment, respectively. $H()$ is the entropy function, and $MI()$ represents mutual information function.

- **Adjusted Rand Index (ARI)** measures similarities between clustering-inference assignment and ground-truth assignment based on consistencies of the two assignment, as follows:

$$ARI = \frac{RI - \mathbb{E}(RI)}{\max(RI) - \mathbb{E}(RI)} \quad (24)$$

where $\mathbb{E}()$ is the expectation operation. RI denotes the rand index computed by the ratio of correct pairwise samples.

- **F1-score (F1)** is defined as the harmonic mean of the precision and the recall of clustering-inference assignment with the following form:

$$F_{\beta} = (1 + \beta^2) \times \frac{P \times R}{(\beta^2 \times P) + R} \quad (25)$$

where β is equal to 1, P denotes the clustering-inference precision, and R represents the clustering-inference recall.

- **Purity** measures consistencies between clustering-inference assignment with ground-truth assignment via the ratio of the number of samples with correct assignment to the number of samples, as follows:

$$\text{Purity}(C, G) = \frac{1}{N} \sum_k \max_j |c_k \cap g_j| \quad (26)$$

where g_j denotes ground-truth class that is the most frequent in the cluster c_k .

5.3 Compared Methods

11 representative multi-view fuzzy clustering methods are selected to validate the performance of DMGFCC by comparison, which can be divided into two groups, i.e., shallow multi-view fuzzy clustering method and deep multi-view fuzzy clustering method. Specifically, the shallow multi-view fuzzy clustering method includes CoFKM [5], WV-Co-FCM [7], CoMK-FC [8], OMVFC-LICAG [15], and DFMKLS [16]. The deep multi-view fuzzy clustering method includes MAGCN [23], DEC [26], IDEC [27], BMVC [28], DEMVC [29], and DSwMFC [25].

5.4 Clustering Results

Table 2 showcases the results of clustering experiments conducted on four multi-view benchmark datasets. In Table 2, DMGFCC on v1 and DMGFCC on v2 represent the clustering results that are achieved by comparing the predicted 10-dimensional one-hot labels on the first view and the second view with the ground-truth ones, respectively.

As shown in Table 2, DMGFCC achieves the state-of-the-art performance in comparison with 11 methods. In detail, DMGFCC attains further comprehensive performance improvements over the high performances of other methods on the five metrics across the four datasets. For example, on the MNIST-USPS dataset, DMGFCC reaches an ACC of 0.9967, a NMI of 0.9909, an ARI of 0.9925, a F1 of 0.9934, and a Purity of 0.9967, surpassing the second-best results by 0.0136, 0.0014, 0.0111, 0.0052 and 0.0127, respectively. Those performance improvements demonstrate that the design of the multi-view Siamese generative adversarial architecture and the generative adversarial dual contrastive clustering loss in DMGFCC enables the state-of-the-art performance that other methods cannot achieve. Meanwhile, DMGFCC on v1 and DMGFCC on v2 report slightly lower clustering performances, which demonstrates that DMGFCC is indeed dependent on the fusion mining of multi-view data patterns. Furthermore, most of the deep multi-view fuzzy clustering methods achieve a better performance than the shallow multi-view fuzzy clustering methods, which demonstrates that the capture of deep relations hidden in multi-view data can benefit the pattern mining of multi-view fuzzy clustering.

Table 2: The clustering results of the compared methods on the four datasets. The best results are highlighted in bold.

Dataset	MNIST-USPS					MNIST-EDGE				
	ACC	NMI	ARI	F1	Purity	ACC	NMI	ARI	F1	Purity
CoFKM [5]	0.2837	0.4391	0.3244	0.4821	0.3848	0.4976	0.2405	0.2487	0.3806	0.2530
WV-Co-FCM [7]	0.5897	0.5662	0.4338	0.4533	0.4880	0.5321	0.5366	0.4821	0.4613	0.6082
CoMK-FC [8]	0.4746	0.4226	0.4117	0.4797	0.5854	0.6338	0.5674	0.5211	0.4709	0.4353
OMVFC-LICAG [15]	0.6372	0.6881	0.6361	0.6572	0.6382	0.7311	0.6456	0.6130	0.7054	0.7541
DFMKLS [16]	0.6787	0.7041	0.6382	0.7238	0.6393	0.7946	0.6467	0.6776	0.7554	0.8328
MAGCN [23]	0.6776	0.5105	0.5015	0.5957	0.4942	0.6938	0.4301	0.4359	0.3588	0.4147
DEC [26]	0.9700	0.9340	0.9416	0.9229	0.9461	0.9194	0.9302	0.9289	0.9117	0.9249
IDEC [27]	0.9038	0.9125	0.9253	0.9407	0.9164	0.9274	0.9345	0.9136	0.9133	0.9279
BMVC [28]	0.9826	0.9865	0.9795	0.9878	0.9728	0.9774	0.9527	0.9661	0.9577	0.9724
DEMVC [29]	0.9822	0.9890	0.9738	0.9293	0.9837	0.9828	0.9705	0.9737	0.9764	0.9732
DSwMFC [25]	0.9831	0.9895	0.9814	0.9882	0.9840	0.9832	0.9708	0.9741	0.9766	0.9790
DMGFCC on v1	0.9825	0.9550	0.9632	0.9657	0.9825	0.9820	0.9518	0.9601	0.9646	0.9820
DMGFCC on v2	0.9769	0.9429	0.9536	0.9550	0.9769	0.8855	0.8520	0.8186	0.8357	0.8855
DMGFCC	0.9967	0.9909	0.9925	0.9934	0.9967	0.9880	0.9739	0.9749	0.9768	0.9880
Dataset	MNIST-INVERSE					EDGE-INVERSE				
	ACC	NMI	ARI	F1	Purity	ACC	NMI	ARI	F1	Purity
CoFKM [5]	0.2707	0.3795	0.2577	0.4393	0.3518	0.2746	0.5352	0.4931	0.4409	0.4448
WV-Co-FCM [7]	0.7719	0.6594	0.5220	0.6102	0.7941	0.7365	0.5363	0.6752	0.6449	0.7903
CoMK-FC [8]	0.6213	0.7952	0.5856	0.6704	0.7620	0.7932	0.6317	0.7687	0.5845	0.5179
OMVFC-LICAG [15]	0.7735	0.7997	0.6989	0.6734	0.7944	0.7978	0.6319	0.7712	0.6982	0.7905
DFMKLS [16]	0.7745	0.8014	0.7031	0.6747	0.7983	0.8016	0.6321	0.7737	0.6993	0.7947
MAGCN [23]	0.6504	0.2925	0.3124	0.2997	0.2487	0.5970	0.2560	0.2262	0.2740	0.2979
DEC [26]	0.9141	0.8878	0.8786	0.8883	0.8886	0.9186	0.8518	0.8510	0.8912	0.9181
IDEC [27]	0.9080	0.9172	0.9197	0.8843	0.8983	0.9108	0.8617	0.8588	0.8749	0.9148
BMVC [28]	0.9273	0.9233	0.9244	0.9309	0.9339	0.9206	0.9333	0.9241	0.9369	0.9286
DEMVC [29]	0.9531	0.9600	0.9614	0.9453	0.9430	0.9869	0.9553	0.9528	0.9454	0.9879
DSwMFC [25]	0.9554	0.9603	0.9629	0.9501	0.9478	0.9899	0.9590	0.9566	0.9460	0.9823
DMGFCC on v1	0.9832	0.9540	0.9628	0.9669	0.9832	0.8873	0.8538	0.8220	0.8380	0.8873
DMGFCC on v2	0.9845	0.9570	0.9657	0.9696	0.9845	0.9819	0.9520	0.9599	0.9647	0.9819
DMGFCC	0.9867	0.9624	0.9703	0.9737	0.9867	0.9906	0.9758	0.9797	0.9814	0.9906

Furthermore, to illustrate the statistical performance of DMGFCC, the Nemenyi test is conducted according to the average ranks of clustering numerical results. Fig. 2 shows the Nemenyi statistical test result of all methods on the four datasets and the five metrics, and there are 2 observations. (1) DMGFCC achieves the first rank of all methods, which demonstrates its comprehensive optimality and stable performance across datasets and metrics, providing rigorous empirical evidence for multi-view clustering method selection. (2) Under the significance level $\alpha = 0.01$ and critical difference $CD = 4.2654$, the outperformance of DMGFCC over most compared methods is statistically significant, which means its undoubtedly optimality in essence rather than random fluctuations. In summary, DMGFCC achieves statistical superiority in the comparison with 11 methods.

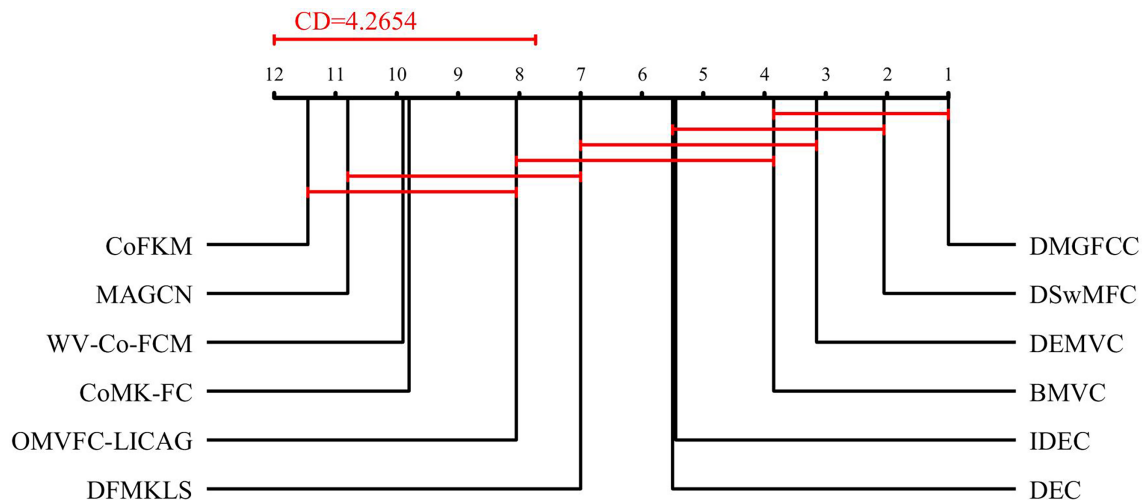


Figure 2: The Nemenyi statistical test result of all methods.

5.5 Visualizing Results

The t-SNE (t-distributed Stochastic Neighbor Embedding) algorithm is conducted on the data points extracted by DMGFCC from four multi-view datasets, to visualize the clustering performance of DMGFCC. The amount of data points used in t-SNE algorithm is 2000 for each dataset, and data points from different ground-truth classes are labeled with different colors.

As shown Fig. 3, the data points, which are endowed with the property of intra-cluster compactness and inter-cluster separation by DMGFCC, lie on the two-dimensional visualization space in a well-classified manner. The data points with the same colors gather while the data points with different colors stay away from each other.

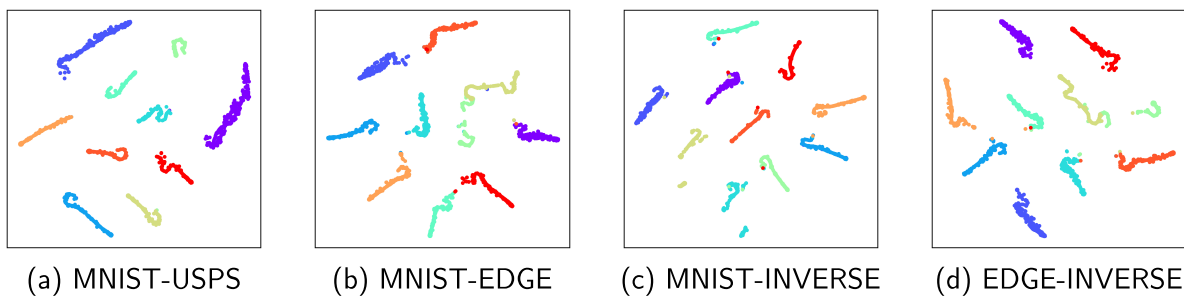


Figure 3: The t-SNE visualizing results of DMGFCC on the four datasets.

In Fig. 4, the confusion matrices on the four multi-view datasets are displayed. In the confusion matrices which are calculated according to the ground-truth labels and the clustering-inference labels, the diagonal elements close to one indicate the good performance on each class. As shown in Fig. 4, all the four confusion matrices are close to the identity matrices with the diagonal elements significantly close to one, which also demonstrates the outperformance of DMGFCC.

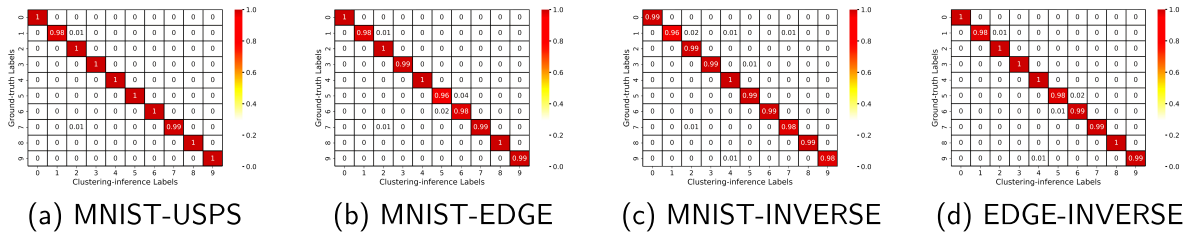


Figure 4: The confusion matrices of DMGFCC on the four datasets.

5.6 Ablation Analysis

To evaluate the contribution of each loss component in DMGFCC, the loss ablation experiments are conducted on the four datasets. As depicted in Table 3, there exist four loss ablation variants where DMGFCC w/o \mathcal{L}_{ga} , DMGFCC w/o \mathcal{L}_{cl} , DMGFCC w/o \mathcal{L}_{mi} , and DMGFCC w/o \mathcal{L}_{fc} denote the removal of the losses \mathcal{L}_{ga} , \mathcal{L}_{cl} , \mathcal{L}_{mi} , and \mathcal{L}_{fc} , respectively. The ACC and NMI results in Table 3 demonstrate that the ablation of each loss component leads to a noticeable degradation on clustering performance on the four datasets, validating the effectiveness of each loss component in DMGFCC.

Table 3: Ablation results on the four datasets in terms of ACC and NMI. The best results are highlighted in bold.

Dataset	MNIST-USPS		MNIST-EDGE		MNIST-INVERSE		EDGE-INVERSE	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
DMGFCC w/o \mathcal{L}_{ga}	0.9668	0.9716	0.8614	0.8769	0.9381	0.8913	0.9221	0.9346
DMGFCC w/o \mathcal{L}_{cl}	0.9171	0.8674	0.9268	0.9327	0.9691	0.9116	0.9747	0.9585
DMGFCC w/o \mathcal{L}_{mi}	0.9824	0.9726	0.8388	0.8882	0.9424	0.9167	0.8704	0.8330
DMGFCC w/o \mathcal{L}_{fc}	0.8764	0.8962	0.8565	0.9425	0.9012	0.8931	0.8573	0.9446
DMGFCC	0.9967	0.9909	0.9880	0.9739	0.9867	0.9624	0.9906	0.9758

5.7 Hyper-Parameter Analysis

To explore the sensitivity of DMGFCC to trade-off hyper-parameters, i.e., λ and γ , the hyper-parameter sensitivity experiment is conducted on the four datasets. In the experiment, the two hyper-parameters are searched in the range of $\{0.2, 0.4, 0.6, 0.8, 1\}$, respectively. As shown in Fig 5, DMGFCC outputs relatively stable ACC results on the four datasets when the two hyper-parameters vary. That is, DMGFCC has a robustness to the selection of values of the trade-off hyper-parameters. Based on this observation, λ and γ are uniformly set as 1 for all datasets.

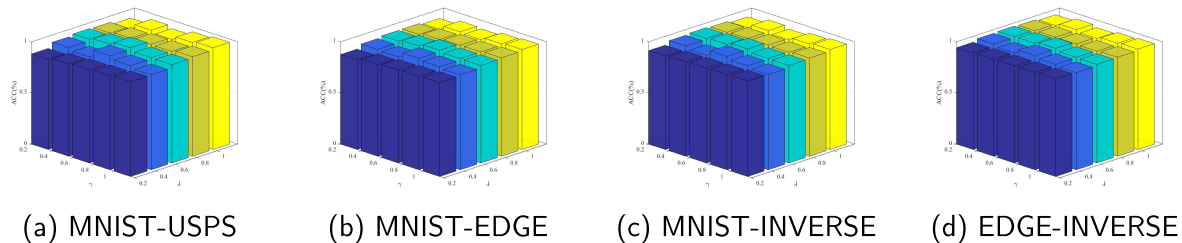


Figure 5: The hyper-parameter sensitivities of DMGFCC on the four datasets.

5.8 Convergence Analysis

To verify the convergence of DMGFCC, Fig. 6 shows the normalized loss curves on the four datasets in 0–80 epochs of training. On the MNIST-USPS dataset, the loss curve decreases rapidly during 0–16 epochs, and then shows slight fluctuations around a loss value of 0.07. On the MNIST-INVERSE dataset, the loss curve has three trends: rapid decrease in 0–13 epochs, slow decrease in 14–37 epochs, and slight fluctuation in the following epochs around the loss value 0.18. And the loss curves on the MNIST-EDGE dataset and the EDGE-INVERSE dataset are almost sandwiched between the two curves above. In general, DMGFCC can converge after 40-epoch training on all four multi-view benchmark datasets, and the different convergence performance may be relevant to their sizes.

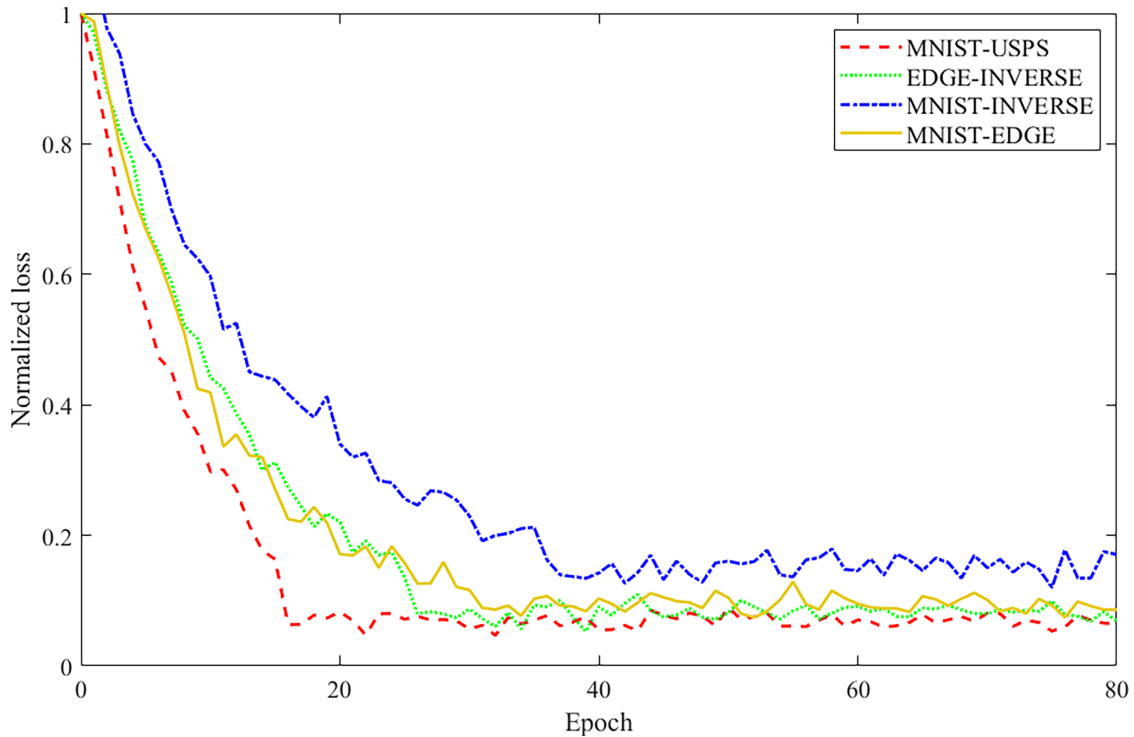


Figure 6: The convergence analysis results of DMGFCC on the four datasets.

5.9 Complexity Analysis

As shown in Algorithm 1, the time complexity and the space complexity of DMGFCC are $\mathcal{O}(nN(VLM^2 + V^2K + VK^2))$ and $\mathcal{O}(VLM^2 + V^2K + VK^2)$, respectively, where n is the number of training epochs, N is the number of multi-view data, V is the number of views, L is the total number of layers in a generator, a discriminator, and a cluster network of each view, M is the maximum dimension of layers, and K is the number of predefined clusters. The detail analyses are shown as follows.

The time complexity. In Algorithm 1, for each training epoch, the first line (Line 1 for short) samples pairwise input vectors for each view, which takes $\mathcal{O}(VM)$ time. Next, Line 2–8 take $\mathcal{O}(NVLM^2)$ time on the forward process of view-specific generator, discriminator, and cluster networks to generate pairwise instances, distinguish real/generated data, and calculate clustering indicator features for each view. Then, Line 9–12 calculates the average of indicator features in each view and find the maximal elements of fusion representations for the clustering assignment, which takes $\mathcal{O}(VK)$ time. In addition, the total clustering loss

is computed in Line 13, which takes $\mathcal{O}(V^2K + VK^2)$ time. In the last three lines of the first loop, i.e., Line 14–16, the backpropagation process is conducted to update parameters in view-specific generator, discriminator, and cluster networks, which takes a similar time to the forward process, i.e., $\mathcal{O}(NVLM^2)$ time. After the training epochs, for each multi-view data x_i , the clustering assignment is obtained by the three lines of the second loop, which takes $\mathcal{O}(VLM^2 + VK)$ time. Considering the two loops, the time complexity of DMGFCC is $\mathcal{O}(nN(VLM^2 + V^2K + VK^2))$.

The space complexity. In the Siamese architecture of DMGFCC, view-specific generator, discriminator, and cluster networks take $\mathcal{O}(VLM^2)$ space. In addition, the sampled pairwise input vectors and fusion representations take $\mathcal{O}(VM)$ and $\mathcal{O}(K)$ space, respectively. Similarly, the computation of total clustering loss takes $\mathcal{O}(V^2K + VK^2)$ space. In summary, the space complexity of DMGFCC is $\mathcal{O}(VLM^2 + V^2K + VK^2)$.

In the experiments, $V = 2$, $L \leq 12$, and $K = 10$. Those values are relatively small compared with M and N , and can be almost negligible. In particular, n is usually much less than N , which is especially advantageous to experiments on large datasets.

6 Conclusion

In this paper, the deep multi-view generative fuzzy contrastive clustering (DMGFCC) is proposed within a Siamese architecture to capture soft patterns of data via clustering-specific fusion representations of invariant structures in informative topologies. Specifically, a multi-view Siamese generative adversarial architecture is designed to capture the joint distribution of data as well as invariant structures, which is composed of the view-specific generator network providing pairwise implicit constraints, the view-specific discriminator network distilling knowledge of real data, and the view-specific cluster network capturing fuzzy patterns of fusion information. Furthermore, a generative adversarial dual contrastive clustering loss consisting of a generative adversarial loss and a dual contrastive clustering loss is devised to supervise the learning of architecture parameters. Finally, experimental results on four benchmark datasets demonstrate the competitive performance of DMGFCC compared with the 11 representative methods. In the future, more multi-view fuzzy clustering schemes will be explored.

Acknowledgement: Not applicable.

Funding Statement: This work was partly supported by the National Natural Science Foundation of China under Grant 62476038.

Author Contributions: The authors confirm contribution to the paper as follows: Conceptualization, data curation, Jing Gao and Peng Li; methodology, software, validation, formal analysis, writing—original draft preparation, visualization, Jianing Zhang; investigation, resources, supervision, project administration, funding acquisition, Zhikui Chen; writing—review and editing, Zhikui Chen, Jing Gao and Peng Li. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: Not applicable.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Xu Y, Liu Y, Shang J, Lin J, Ma D. OceanAgent: a small-scale multi-modal assistant for ocean exploration. *Expert Syst Appl.* 2026;298:129640.
2. Lei L, Huang J, Zhou Y. Multimodal fusion-based spatiotemporal incremental learning for ocean environment perception under sparse observation. *Inf Fusion.* 2024;108(7):102360. doi:10.1016/j.inffus.2024.102360.
3. Li M, Hou Y, Song X, Hou C, Xiong Z, Ma D. Self-attention-guided multiindicator retrieval for ocean surface wind field with multimodal data augmentation and fusion. *IEEE Trans Geosci Remote Sens.* 2024;62(13):1–22. doi:10.1109/tgrs.2024.3452136.
4. Bai L, Qiu L, Zheng J, Zhang Y, Chen X, Sun Y. A parallel convolution attention and temporal sequence attention neural network approach for ocean current prediction incorporating spatial-temporal coupling mechanism. *Expert Syst Appl.* 2025;281(2):127681. doi:10.1016/j.eswa.2025.127681.
5. Cleuziou G, Exbrayat M, Martin L, Sublemontier JH. CoFKM: a centralized method for multiple-view clustering. In: *Proceedings of the 2009 Ninth IEEE International Conference on Data Mining; 2009 Dec 6–9; Miami, FL, USA.* p. 752–7.
6. Yang H, Deng Z, Zhang W, Wu Q, Choi K, Wang S. End-to-end multiview fuzzy clustering with double representation learning and visible-hidden view cooperation. *Trans Fuzzy Syst.* 2024;32(2):483–97. doi:10.1109/TFUZZ.2023.3300925.
7. Jiang Y, Chung F, Wang S, Deng Z, Wang J, Qian P. Collaborative fuzzy clustering from multiple weighted views. *IEEE Trans Cybern.* 2015;45(4):688–701. doi:10.1109/TCYB.2014.2334595.
8. Zeng S, Wang X, Cui H, Zheng C, Feng DD. A unified collaborative multikernel fuzzy clustering for multiview data. *IEEE Trans Fuzzy Syst.* 2018;26(3):1671–87. doi:10.1109/TFUZZ.2017.2743679.
9. Kumar D, Maji P. Discriminative deep canonical correlation analysis for multi-view data. *IEEE Trans Neural Netw Learn Syst.* 2024;35(10):14288–300. doi:10.1109/TNNLS.2023.3277633.
10. Yu X, Jiang Y, Chao G, Chu D. Deep contrastive multi-view subspace clustering with representation and cluster interactive learning. *IEEE Trans Knowl Data Eng.* 2025;37(1):188–99. doi:10.1109/TKDE.2024.3484161.
11. Che H, Li C, Leung M, Ouyang D, Dai X, Wen S. Robust hypergraph regularized deep non-negative matrix factorization for multi-view clustering. *IEEE Trans Emerg Top Comput Intell.* 2025;9(2):1817–29. doi:10.1109/TETCI.2024.3451352.
12. Zhao M, Yang W, Nie F. Deep multi-view spectral clustering via ensemble. *Pattern Recognit.* 2023;144(10):109836. doi:10.1016/j.patcog.2023.109836.
13. Wang Y, Chen L. Multi-view fuzzy clustering with minimax optimization for effective clustering of data from multiple sources. *Expert Syst Appl.* 2017;72(4):457–66. doi:10.1016/j.eswa.2016.10.006.
14. Yang M, Sinaga KP. Collaborative feature-weighted multi-view fuzzy c-means clustering. *Pattern Recognit.* 2021;119:108064. doi:10.1016/j.patcog.2021.108064.
15. Zhang C, Chen L, Shi Z, Ding W. Latent information-guided one-step multi-view fuzzy clustering based on cross-view anchor graph. *Inf Fusion.* 2024;102(6):102025. doi:10.1016/j.inffus.2023.102025.
16. Yin J, Sun S, Wei L, Wang P. Discriminatively fuzzy multi-view k-means clustering with local structure preserving. In: *Proceedings of the AAAI'24: AAAI Conference on Artificial Intelligence; 2024 Feb 20–27; Vancouver, BC, Canada.* p. 16478–485.
17. Tzortzis G, Likas A. Kernel-based weighted multi-view clustering. In: *Proceedings of the 2012 IEEE 12th International Conference on Data Mining; 2012 Dec 10–13; Brussels, Belgium.* p. 675–84.
18. Guo D, Zhang J, Liu X, Cui Y, Zhao C. Multiple kernel learning based multi-view spectral clustering. In: *Proceedings of the 2014 22nd International Conference on Pattern Recognition; 2014 Aug 24–28; Stockholm, Sweden.* p. 3774–9.
19. Ye Y, Liu X, Yin J, Zhu E. Co-regularized kernel k-means for multi-view clustering. In: *Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR); 2016 Dec 4–8; Cancun, Mexico.* p. 1583–8.
20. Trosten DJ, Løkse S, Jenssen R, Kampffmeyer M. Reconsidering representation alignment for multi-view clustering. In: *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2021 Jun 20–25; Nashville, TN, USA.* p. 1255–65.

21. Gao Q, Lian H, Wang Q, Sun G. Cross-modal subspace clustering via deep canonical correlation analysis. *Proc AAAI Conf Artif Intell.* 2020;34(4):3938–45. doi:10.1609/aaai.v34i04.5808.
22. Mao Y, Yan X, Guo Q, Ye Y. Deep mutual information maximin for cross-modal clustering. *Proc AAAI Conf Artif Intell.* 2021;35(10):8893–901. doi:10.1609/aaai.v35i10.17076.
23. Cheng J, Wang Q, Tao Z, Xie D, Gao Q. Multi-view attribute graph convolution networks for clustering. In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20)*; 2021 Jan 7–15; Yokohama, Japan. p. 2973–9.
24. Yin M, Huang W, Gao J. Shared generative latent representation learning for multi-view clustering. *Proc AAAI Conf Artif Intell.* 2020;34(4):6688–95. doi:10.1609/aaai.v34i04.6146.
25. Shi M, Zhao X, Yin X, Xiao Y, Guo J. Deep self-weighted multi-view fuzzy clustering. *Knowl Based Syst.* 2025;328(2):114158. doi:10.1016/j.knosys.2025.114158.
26. Xie J, Girshick RB, Farhadi A. Unsupervised deep embedding for clustering analysis. In: *Proceedings of the 33rd International Conference on International Conference on Machine Learning*; 2016 Jun 19–24; New York, NY, USA. p. 478–87.
27. Guo X, Gao L, Liu X, Yin J. Improved deep embedded clustering with local structure preservation. In: *IJCAI'17: Proceedings of the 26th International Joint Conference on Artificial Intelligence*; 2017 Aug 19–25; Melbourne, VIC, Australia. p. 1753–9.
28. Zhang Z, Liu L, Shen F, Shen HT, Shao L. Binary multi-view clustering. *IEEE Trans Pattern Anal Mach Intell.* 2019;41(7):1774–82. doi:10.1109/tpami.2018.2847335.
29. Xu J, Ren Y, Li G, Pan L, Zhu C, Xu Z. Deep embedded multi-view clustering with collaborative training. *Inf Sci.* 2021;573:279–90. doi:10.1016/j.ins.2020.12.073.